Summer School on Operational Research and Applications

May 22 – 26, 2016

National Research University Higher School of Economics,

Nizhny Novgorod

Faculty of Informatics, Mathematics and Computer Science

Laboratory of Algorithms and technologies for network analysis

Sunday, May 22

Arrival of participants

Monday, May 23

Room 209 HSE, 136 Rodionova Str.

9:30-10:00 Registration of participants

10:00 -10:50 Panos Pardalos

Opening lecture: Network Robustness from an Information Theory Perspective

10:50 -11:10 Coffee break

11:10 -12:00 Panos Pardalos

Lecture: Computational models and challenging optimization problems

12:10 -13:00 Oleg Prokopyev

Lecture 1: Finding Influential (Central) Groups in Networks using Betweenness Centrality

13:00 -14:30 Lunch

14:30 -15:20 Oleg Prokopyev

Lecture 2: Finding Influential (Central) Groups in Networks using Betweenness Centrality

15:30 -16:20 Valery Kalyagin

Lecture: Network models for stock market

16:20 -16:30 Coffee break

16:30 -17:30 Student session (each talk is about 10 min):

Tamara Shcheglova *Socio-political "open pages" classification in social networks*

Nikita Putikhin *Usage of continuous formulation of boolean expressions in the stochastic local search algorithm for boolean satisfiability problem*

Nikolay Korolkov *Game theory application in strategic bank management*

Alexander Gnusarev *Comparison of the two heuristic algorithms for the location and design problem*

Vladimir Ivashkin *Comparison of graph measures in the context of clustering*

Room 207 HSE, 136 Rodionova Str.

17:30 -18:30 Scientific evening with company Itseez (in Russian):

- Kirill Kornyakov, R&D Director at Itseez: presentation of the company and its projects, talk on computer vision
- Ilya Lysenkov, Technical Director at itSeez3D: presentation of itSeez3D project, demonstration of the 3D-scanner
- Alexey Gruzdev, Software engineer: HSE graduated student's career
- Boris Yastrebov, HR director: Itseez summer programs for students
- Q&A + pizza and soft drinks (room 401)

Tuesday, May 24

Room 209 HSE, 136 Rodionova Str.

10:00-10:50 René van Bevern

Lecture 1: Fixed-parameter algorithms: circumventing intractability by exploiting input structure

10:50 -11:10 Coffee break

11:10-12:00 René van Bevern

Lecture 2: Problem kernelization: polynomial-time data reduction with provable effect

12:10-13:00 Dmitry Ignatov

Lecture 1: Pattern Mining and Multi-modal Clustering: searching for optimal patterns

13:00-14:30 Lunch

14:30-15:20 Dmitry Ignatov

Lecture 2: Pattern Mining and Multi-modal Clustering: searching for optimal patterns

15:30-16:20 René van Bevern

Seminar: Fixed-parameter linear-time algorithms for finding Colorful Independent Sets

16:20 -16:40 Coffee break

16:40 -17:30 Dmitry Ignatov

Seminar: Practical Tools for Pattern Mining and Multi-modal Clustering

Wednesday, May 25

Room 209 HSE, 136 Rodionova Str.

10:00-10:50 Nikolay Zolotykh

Lecture: Polyhedral computations

10:50 -11:20 Coffee break

11:10-12:00 Oleg Prokopyev

Seminar: Integer Programming for Finding Maximum Quasi-Cliques and Dense Subgraphs

12:10-13:00 Sergey Nikolenko

Lecture 1: Competitive Analysis in Buffer Management

13:00 – 14:00 Lunch

14:30-15:20 Sergey Nikolenko

Lecture 2: Competitive Analysis in Buffer Management

15:30-16:20 Sergey Nikolenko

Seminar: Competitive Analysis in Buffer Management

16:20 -16:40 Closing

Thursday, May 26

Departure of participants

Network Robustness from an Information Theory Perspective

Panos M. Pardalos

University of Florida, USA & National Research University Higher School of Economics, Russia

A crucial challenge in network theory is the study of the robustness of a network when facing a sequence of failures. We propose a novel methodology to measure the robustness of a network to component failures or targeted attacks based on Information Theory, that considers measurements of the structural changes caused by failures of the network's components providing a dynamical information about the topological damage. The methodology is comprehensive enough to be used with different probability distributions and provides a dynamic profile that shows the response of the network's topology to each event, quantifying the vulnerability of these intermediate topologies.

Finding Influential (Central) Groups in Networks using Betweenness Centrality

Oleg Prokopyev

University of Pittsburgh, USA

Graph-theoretical models arise in a variety of application areas due to their elegance and inherent ability to logically represent (as edges) important relationships, e.g., communication, social and transportation links, between structural elements (i.e., nodes) of complex systems. The notion of betweenness centrality allows us to rank all nodes (or groups of nodes) according to their significance to the network structure under the assumption that all node-to-node communications are performed through shortest paths. Besides social networks, betweenness centrality has found a number of other application domains, both as a direct tool for network analysis or as an important building block within various types of network algorithms. In this talk we consider the problem of identifying the most influential (or central) group of nodes (of some predefined size) in a network by assuming that such group has the largest value of betweenness centrality or one of its variants, e.g., the length-scaled or the bounded-distance betweenness centralities. We demonstrate that this problem can be modeled as a mixed integer program (MIP) that can be solved for reasonably sized network instances using off-the-shelf MIP solvers. We also discuss interesting relations between the group betweenness and the bounded-distance betweenness centrality concepts. In particular, we exploit these relations in an algorithmic scheme to identify approximate solutions for the original problem. Furthermore, we generalize our approach for identification of not only the most central groups of nodes, but also central general groups of graph elements (either nodes or edges exclusively, or their combination according to some pre-specified criteria) satisfying, if necessary, some additional cohesiveness properties (e.g., the targeted group should form a clique or a k-club). We conduct extensive computational experiments with different types of real-life and synthetic network instances to show the effectiveness and flexibility of the proposed optimization framework. Our experiments reveal some interesting insights into the properties of influential groups of graph elements modeled using the maximum betweenness centrality concept or one of its variations.

Network models for stock market

Valery Kalyagin

National Research University Higher School of Economics, Russia

Stock market can be modeled as a random variables network: nodes of network are random variables with joint distribution from some class of multivariate distributions, weighted links of network are given by some measure of association (dependence or similarity) between random variables. Network structure is a subgraph in this complete weighted graph. Examples of structures are: maximum spanning tree, planar maximally filtered graph, threshold networks, threshold (market) graph, cliques and independent sets in threshold graph and so on. The main problem is to identify a network structure from observations. The following topics will be covered in the lecture

- Uncertainty of statistical procedures for network structure identification. Comparison of uncertainty for different network structures and different markets.

- Comparison of different market network models. Robustness of statistical procedures for network structures identification.

Socio-political “open pages” classification in social networks

Tamara Shcheglova

National Research University Higher School of Economics, International Laboratory for Applied Network Analysis, Russia

Facebook founder eventually approached his goal: the world has truly become more open and connected. There are currently interactions even between strangers on Facebook and other social networks, who communicate not only in the areas of interest but in the areas of broad socio-political concern. The relationship between Ukraine and Russia constitute the hottest issue today. This research is an attempt to highlight the issues of the information campaign in Ukraine which is obviously of growing scientific interest. Socio-political “open pages” grouping on the basis of linguistic and network grounds provides understanding of the mechanisms of work of these pages and thus the perception of the situation by users. The aim of the work is the comparison of network and linguistic bases for grouping socio-political "open pages" relating to the events in Ukraine in social networks. As a result of this work the data set of socio-political "open pages" Vkontakte was divided into two clusters in several ways, and the key actors of these pages have been identified. On the basis of the study it can be argued that the network clustering allows to separate the data set into clusters more clearly, taking into account the real connections between open pages.

Usage of continuous formulation of boolean expressions in the stochastic local search algorithm for boolean satisfiability problem

Nikita Putikhin

National Research University Higher School of Economics, Faculty of Informatics, Mathematics and Computer Science, Russia

The boolean satisfiability problem (SAT) is one of the most fundamental problems in Computer Science. It is formulated as a question: “is there such an assignment of variables that satisfies given boolean formula?”. We examine two main approaches used in SAT solving: complete DPLL-based algorithms and stochastic local search (SLS) algorithms of WalkSAT family. While DPLL does not benefit from the usage of continuous formulations, for WalkSAT it may provide better variable selection strategy. We present the modification of gNovelty+ SLS algorithm to use continuous form of boolean variables.

Game theory application in strategic bank management

Nikolay Korolkov

National Research University Higher School of Economics, Faculty of Economics, Russia

This paper attempts to show that banks can use instruments of the game theory in their strategic bank management. In this study, we implemented simulation game models of the process of interaction between the bank ("Ассоциация") and its competitors, resulting in the construction of a competitive strategy based on the principle of maximum guaranteed result that allows to respond effectively to changes in competitors' pricing policy. The main contribution of this study is a developed and experimentally tested method for constructing the competitive strategies of commercial bank by a set of tools, such as the game theory, cluster analysis and econometric modeling.

Comparison of the two heuristic algorithms for the location and design problem

Alexander Gnusarev

Sobolev Institute of Mathematics, Omsk Branch, Omsk, Russia

This paper is devoted to location and design problem. Two competing persons seeking to maximize volume of customers demand. One of the participants has already opened his own facilities. Another person is required to determine the place and the options of opening new facilities in order to attract the largest share of total demand. The peculiarity of the problem lies in the fact that the share of demand served by the company is not fixed and depends on the location and type of business. We construct Variable neighborhoods descent algorithm and compared it with greedy weight heuristic. Experimental tuning of the parameters of both algorithms was carried out. Computational experiments were performed on special test instances.

Comparison of graph measures in the context of clustering

Vladimir Ivashkin

Moscow Institute of Physics and Technology, Department of Radio Engineering and Cybernetics, Russia

In this work we address the problem of finding the best parametric family of graph proximity/dissimilarity measures and the best family parameters in the clustering problems for graph nodes. We carry out a number of experiments on the comparison of measures for both random clustered networks and well-known datasets. During the experiments, we pay attention not only to the maximum quality reached by each family, but also to their robustness with respect to the family parameter. In our studies we found that taking elementwise logarithm for some measures significantly improves the quality of clustering. This conclusion is confirmed by studying inter-cluster and intra-cluster distances in the graph.

Fixed-parameter algorithms: circumventing intractability by exploiting input structure

René van Bevern

Novosibirsk State University, Russia

This lecture introduces basic concepts and techniques for developing fixed-parameter algorithms, a recent approach to optimally solving NP-hard problems. NP-hard problems presumably cannot be solved optimally within a worst-case running time that grows polynomially with the input size. Yet it is often possible to derive so-called fixed-parameter algorithms, whose running time grows polynomially or only linearly with the input size and exponentially only with respect to other instance parameters. Thus, fixed-parameter algorithms lead to efficient algorithms for NP-hard problems in applications where these parameters are small, for example, logarithmic in the input size.

Problem kernelization: polynomial-time data reduction with provable effect

René van Bevern

Novosibirsk State University, Russia

Data reduction is an important tool when solving problems on large input data. In general, it is difficult to derive data reduction algorithms that are fast, correct, and effective at the same time: for example, consider a polynomial-time data reduction algorithm for an NP-hard problem that does not change the value of an optimal solution but provably reduces the input size by at least one bit. Using repeated application, this algorithm would solve the NP-hard problem in polynomial time, thus implying P=NP. The field of fixed-parameter algorithms spawned the concept of problem kernelization: provably correct and effective polynomial-time data reduction. The main idea is to bound the size of the reduced instance by a function not of the input size, but of additional instance parameters, so that repeated application does not necessarily shrink the instance further. This lecture gives an introduction and various examples for problem kernelization algorithms.

Pattern Mining and Multi-modal Clustering: searching for optimal patterns/ Practical Tools for Pattern Mining and Multi-modal Clustering

Dmitry Ignatov

National Research University Higher School of Economics, Russia

In these lectures, we will examine basic models and algorithms of Pattern Mining, a crucial part of modern Data Mining. This discipline aims at finding novel, useful, and interpretable regularities in large amounts of data. One the most typical examples is so called Frequent Itemset Mining (FIM) that grew up in early 90s from market basket analysis. It also includes Association Rules Mining, i.e. usage of rules in form A→B, meaning, for example, that people who bought A have also bought B (according to the input data) with some level of confidence. Nowadays these approaches and algorithms were extended and used in many other applications like web usage mining, information retrieval, web advertising, resource sharing systems, recommender systems, market segmentation, near duplicate detection, gene expression analysis, communities detection in SNA, learning and game analytics, and many others. Clustering, classification and outlier detection are also one of the related topics with respect to applications of Pattern Mining. We will go through basic algorithms like Apriori and FP-growth, talk about sequential patterns with examples from demography and mention graph pattern mining. We will also discuss compact ways of frequent itemsets representation like closed and maximal closed frequent itemsets as well as interestingness measures to extract the most (potentially) useful patterns. As mathematical basis for our treatment we will use Formal Concept Analysis (FCA), an applied branch of modern lattice theory suited for Data Analysis. It is important to note that FCA was proposed almost 10 years prior FIM and based on the notions of Galois connection, closure operator, and so called concept lattice. In the second part of the talk, we will introduce the notions of bicluster and tricluster as examples of patterns from multi-modal clustering; this is an extension of Frequent Itemset Mining and Formal Concept Analysis, and it again can be considered as a part of Pattern Mining. In the practical part, we will experiment with Frequent Itemset Mining and Formal Concept Analysis tools such as SPMF, Orange 2.7, Concept Explorer and several others.

Fixed-parameter linear-time algorithms for finding Colorful Independent Sets

René van Bevern

Novosibirsk State University, Russia

In this seminar, we will study an NP-hard graph problem that naturally arises in scheduling: given a vertex-colored interval graph, find a maximum-cardinality independent set whose vertices have pairwise distinct colors. Using the techniques presented in the lectures on fixed-parameter algorithms and kernelization, we will develop linear-time data reduction algorithms and, moreover, algorithms for solving this generally NP-hard problem in linear time if the number of colors in the input graphs is constant.

Polyhedral computations

Nikolay Zolotykh

National Research Lobachevsky State University of Nizhni Novgorod, Russia

We consider different computational problems dealing with convex polyhedra in general dimension. First of all we focus on the representation conversion problem (between halfspace and vertex representations) and discuss some algorithms. Also, some applications of polyhedral computation (in optimization, computational biology, program verification etc.) will be presented.

Integer Programming for Finding Maximum Quasi-Cliques and Dense Subgraphs

Oleg Prokopyev

University of Pittsburgh, USA

Given a simple graph and a constant gamma, a gamma-quasi-clique is defined as a subset of vertices that induces a subgraph with an edge density of at least gamma. This well-known clique relaxation model arises in a variety of application domains. The maximum gamma-quasi-clique problem is to find a gamma-quasi-clique of maximum cardinality in the graph and is known to be NP-hard. This talk discusses mixed integer programming formulations for solving the maximum gamma-quasi-clique problem. The corresponding linear programming relaxations are also analyzed. Finally, we also consider generalizations for solving the maximum f-dense subgraph problem, which, for a given function f seeks for the largest k such that there is a subgraph induced by k vertices with at least f (k) edges.

Competitive Analysis in Buffer Management

Sergey Nikolenko

St. Petersburg Department of the Steklov Mathematical Institute, Russia

An important new direction of research in buffer management, gaining popularity over the last decade, deals with worst-case bounds on the competitive ratios of buffer management policies in various settings. This is a young field of study, and many cutting edge problems are easy to understand and, perhaps, quite possible to solve. In the short course, we will review the main definitions of competitive analysis, prove several worst-case bounds for illustration, and survey the main problem settings and the current frontier of the field.