8th International Conference on Network Analysis NET 2018

Moscow, May 18-19, 2018

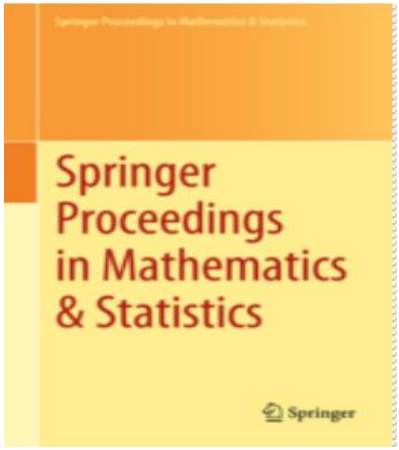
Laboratory of Algorithms and Technologies for Network Analysis
National Research University
Higher School of Economics, Nizhny Novgorod
Laboratory of Advanced Combinatorics and Network Applications,
National Research University
Moscow Institute of Physics and Technology

RUSSIAN SCIENCE FOUNDATION

NATIONAL RESEARCH UNIVERSITY

MIPT

Yandex



List of plenary talks

Vladimir Boginski, Associate Professor, Industrial Engineering & Management Systems University of Central Florida: "Robustness and Vulnerability of Interdependent Infrastructure Networks: Mathematical Modeling and Optimization Aspects"

Ernesto Estrada, Chair of Mathematics, Chair in Complexity Sciences, Department of Mathematics and Statistics, University of Strathclyde Glasgow, UK: "Machine Learning Analysis of Complex Networks in Hyperspherical Space"

Mikhail Isaev, Lecturer, School of Mathematical Sciences, Monash University Australia and MIPT: "Advances in the enumeration of regular graphs and not-so-regular graphs"

Marcello Pelillo, Professor of Computer Science, Ca' Foscari University of Venice, Director of European Centre for Living Technology (ECLT), Head of the Computer Vision and Pattern Recognition Lab DAIS: "Grouping Games: Finding Clusters in Graphs, Digraphs and Hypergraphs"

Martín Gómez Ravetti, Federal University of Minas Gerais, Brazil: "Quantification of graph dissimilarities using Information Theory"

Angelo Sifaleras, School of Information Sciences, University of Macedonia, Greece: 'Recent applications of Variable Neighborhood Search".

Tutorial for young scientists

Theodore Trafalis, University of Oklahoma, USA: "Max-Min Fairness and its Applications".

Location: Moscow, Yandex office, Lva Tolstogo street, 16

Friday, May 18

09:00 – 09:30 Registration

09:30 – 09:50 Panos M. Pardalos

Networks: History, The Present, and Future Challenges

09:50 – 10:40 Marcello Pelillo,

Grouping Games: Finding Clusters in Graphs, Digraphs and Hypergraphs

10:40 – 10:50 Coffee Break

10:50 – 11:40 Martín Gómez Ravetti,

Quantification of graph dissimilarities using Information Theory

11:40 – 12:40 Session 1

Marco Fiorucci - *Graph Summarization Using Regular Partitions*
Sergey Makrushin - *Network structure and scheme analysis of the Russian-language segment of Wikipedia*
Sergey Shvydun - *Computational Complexity of SRIC and LRIC indices*
Alexander Veremyev - *Integer programming techniques for modeling clique relaxations in networks*

12:40 – 14:00 Lunch Break

14:00 – 14:50 Angelo Sifaleras,

Recent applications of Variable Neighborhood Search.

14:50 – 15:00 Coffee Break

15:00 – 16:30 Session 2

Valentina Kuskova - *Individual project effectiveness: individual network predictors*
Galina Gradoselskaya - *Mapping of politically active groups on social networks of Russian regions (on the example of Karachay-Cherkess Republic)*
Ilya Makarov - *Co-authorship recommender System and Graph Embeddings*
Tamara Shcheglova - *Differentiation of discursive political space in social networks*
Dmitry Zaytsev - *Advisors Role in Russian Education Policy: Application of Network Analysis*
Nikolai Katargin - *Optimization of Network Schedule of Set of Operations in Excel*

16:30 – 16:40 Coffee Break

16:40 – 18:00 **Tutorial for young scientists:** Theodore Trafalis

Max-Min Fairness and its Applications

Saturday, May 19

09:30 – 10:20 Ernesto Estrada,

Machine Learning Analysis of Complex Networks in Hyperspherical Space

10:20 – 10:40 Coffee Break

10:40 – 11:30 Mikhail Isaev,

Advances in the enumeration of regular graphs and not-so-regular graphs

11:30 – 12:30 Session 3

Mikhail Chernoskutov - *Large-Scale Graph Processing Systems for Network Science Applications*

Alexander Ivanov - *Algorithms of Extracting Patent-Related Data from Professional Blogs to Improve Patent Value Prediction*

Pierre Miasnikof - *Metaheuristcs for Large-Scale Graph Clustering*

Alexander Rubchinsky - *Dichotomy Complexity of Graphs: Notions and Algorithms*

12:30 – 14:00 Lunch Break

14:00 – 14:50 Vladimir Boginski,

Robustness and Vulnerability of Interdependent Infrastructure Networks: Mathematical Modeling and Optimization Aspects

14:50 – 16:20 Session 4

Dmitry Gribanov - *FPT-algorithms for some problems related to integer programming*

Petr Koldanov - *On equivalence of network structures in different random variables networks*

Anvar Kurmukov - *Continuous connectome model of brain network exhibits scale-free property*

Natalia Meshcheryakova - *Indirect influence assessment in the context of retail food network*

Alexander Ponomarenko - *A Method for Overlapping Community Detection in Networks based on Link Partitioning and Partitioning Around Medoids*

Oleg A. Prokopyev - *On maximum degree-based quasi-clique problem*

16:20 – 16:30 Coffee Break

16:30 – 18:00 Session 5

Ilya Bychkov - *From Cell Formation Problem to Biclustering and Graph Editing*

Mikhail Batsyn - *On NP-hardness of the Cell Formation Problem*

Ekaterina N. Beresneva - *A survey of existing approaches to the Capacitated Vehicle Routing Problem*

Mariia Gordenko - *The Generalized Traveling Salesman Problem*

Oxana Mikhailova - *Social Mechanisms of the Subject Area Formation. The Case of "Digital Economy"*

Dmitry Sirotkin - *On local graph transformations for the vertex k-coloring problem*

Valery Kalyagin - *Loss function, unbiasedness, and optimality of Gaussian graphical model selection*

On NP-hardness of the Cell Formation Problem

Mikhail Batsyn

National Research University Higher School of Economics
https://nnov.hse.ru/en/latna/

In this talk we consider the Cell Formation Problem which consists in optimal biclustering of machines together with parts processed on them into production cells. We show the connections of this problem with Minimum Cut and Bicluster Graph Editing Problems. The proof of NP-hardness for the Bicluster Graph Editing Problem belonging to Amit (2004) is presented. It provides the NP-hardness of one of the Cell Formation Problem formulations. The NP status of other formulations is also discussed in this talk.

A survey of existing approaches to the Capacitated Vehicle Routing Problem

Ekaterina N. Beresneva

National Research University Higher School of Economics

Vehicle Routing Problem (VRP) is one of the most widely known questions in a class of combinatorial optimization problems. It is concerned with the optimal design of routes to be used by a fleet of vehicles to serve a set of customers. In this study we analyze Capacitated Vehicle Routing Problem (CVRP) – a subcase of VRP, where the vehicles have a limited capacity. CVRP is aimed at savings in the global transportation costs. Typical applications of CVRP are delivery of goods, solid waste collection, street cleaning etc. The problem is NP-hard, therefore heuristic algorithms which provide near-optimal polynomial-time solutions will be considered instead of the exact ones. The aim of this article is to present a new adaptation of mathematical formulations of CVRP and to make a survey on methods for solving each type of this problem. This paper provides an overview of the most perspective methods to be realized in later works.

Robustness and Vulnerability of Interdependent Infrastructure Networks: Mathematical Modeling and Optimization Aspects

Vladimir Boginski

Associate Professor of Industrial Engineering and Management Systems at the University of Central Florida (Orlando, FL)

Interdependent networks arise in many application domains associated with infrastructure systems, such as electric power, telecommunication, and transportation networks. In a real-world setup, these systems interact with each other, so that disruptions/failures of components in one of the systems may affect the performance of other systems that depend on those components. Thus, failures can propagate through interdependent networked systems in a cascading fashion, where a failure of a component in one system may cause a failure of multiple components in another system, and so on. Important research issues that need to be addressed in studying such interdependent networks include developing mathematical models of cascading failures, as well as assessing robustness/vulnerability of these networks via appropriate quantitative metrics. Furthermore, these mathematical representations may allow one to formulate and solve optimization problems that can potentially reveal interesting properties of the underlying systems and optimal strategies for enhancing their resilience. In this presentation, we will discuss some of our recent results and identify challenges and potential future research directions in this area.

From Cell Formation Problem to Biclustering and Graph Editing

Ilya Bychkov

National Research University Higher School of Economics
https://nnov.hse.ru/en/latna/

The Cell Formation Problem (CFP) has been studied as an optimization problem in manufacturing for more than 90 years. It consists of grouping machines and parts into manufacturing cells in order to maximize loading of cells and minimize movement of parts from one cell to another. In this talk we will consider the classical CFP formulation with different objective functions from literature. Also, we will show the connection between CFP some other interesting problems such as biclustering problems, minimum k-cut problem and bicluster graph editing problem. An effective integer linear programming model for solving CFP will be presented. The performance of the model will be demonstrated on two datasets. The first one is the classical 35GT problems dataset. The second one is a new dataset which contains less popular CFP instances from the literature. The suggested model has obtained optimal solutions have been found for 63 of the 67 considered problem instances and several new solutions unknown before have been obtained. The computational times are greatly decreased comparing to the state-of-art approaches.

Large-Scale Graph Processing Systems for Network Science Applications

Mikhail Chernoskutov

Krasovskii Institute of Mathematics and Mechanics

Nowadays, supercomputing research community are aware about a lot of big graphs processing systems. Systems like Parrallel BGL, Pregel, Giraph are widespread now. It is interesting that all of these systems are differs a lot in terms of (typical) problem size, degree of available parallelism, set of implemented algorithms and a lot of other features. Hence, the presentation is dedicated to analysis of these (open) graph processing systems as well as important question of searching an “ideal” graph processing system, which can address all required user tasks in most efficient way.

Machine Learning Analysis of Complex Networks in Hyperspherical Space

Ernesto Estrada

Chair of Mathematics, Chair in Complexity Sciences,
Department of Mathematics and Statistics, University of Strathclyde
Glasgow, U.K

Complex networks represent the topological skeleton of complex systems in a variety of scenarios, ranging from molecular to social and ecological ones. We discuss here the way in which information diffuses through the nodes and edges of such networks in an “all-routes” way instead of by using “shortest-paths” only. Then, we defined a communicability function that accounts for such kind of all-routes communication in networks. We then prove analytically that such communicability function induces an embedding of any network into a hyperspherical space. The main parameters for defining this embedding are the communicability distance and communicability angles. We then use machine learning techniques, such as nonmetric multidimensional scaling (NMDS) and clustering analysis to extract information about these networks in the (n-1)-spheres in which they “live”. Using NMDS we produce pictorial representations of random and real-world networks as 3-dimensional Euclidean spheres. The clustering analysis by using K-Means reveals the existence of communities in real-world networks which coincide with the ground truth structure of such networks. We finally illustrate some examples of the importance of considering communicability shortest paths between nodes as a way in which items are transmitted through real-world networks.

Graph Summarization Using Regular Partitions

Francesco Pelosin[1], **Marco Fiorucci[1]**, and Marcello Pelillo[1,2]

[1] DAIS, Ca Foscari University, Via Torino 155, Venice, Italy
[2] ECLT, Ca Foscari University, S. Marco 2940, Venice, Italy

The world we live in is becoming more and more interconnected and huge amounts of data are produced and stored every day by different interrelated entities. This high-throughput generation calls for the development of efficient algorithms to understand and process large and noisy network data. To address this challenging task, we develop a principled approach to summarize large graphs based on regular partitions, the existence of which was first established in a celebrated result proved by Endre Szemeredi in the mid-1970s [3]. A regular partition is defined as a partition of the vertex set into a bounded number of random-like bipartite graphs, called regular pairs. In particular, a regular pair is a highly uniform bipartite graph in which the density of any reasonably sized subgraphs is about the same as the overall density of the bipartite graph. Although polynomial-time algorithms are available for finding regular partitions, their practical applicability is hindered by the presence of astronomically large hidden constants.

In this talk, we will present a heuristic for finding approximately regular partitions, based on a well-known O(n2:376) exact algorithm of Alon et al. [1]. This allows us to summarize an input graph by constructing the so-called reduced graph, whose vertices are the sets of the regular partition, and where two vertices are connected by an edge if the corresponding sets form a regular pair with edge density above a given threshold. The novelty of our approach is that the regularity lemma provides us a principled way to separate structure from noise. This is guaranteed by the so-called Key Lemma [2] which states that, under certain conditions, the reduced graph inherits many of the essential structural properties of the original graph. Preliminary experiments provide evidence of the effectiveness and the robustness of the proposed framework, especially when working with noisy data.

References

1. Alon, N., Duke, R., Lefmann, H., Rodl, V., Yuster, R.: The algorithmic aspects of the regularity lemma. Journal of Algorithms 16(1), 80 { 109 (1994)

2. Komlos, J., Shokoufandeh, A., Simonovits, M., Szemeredi, E.: The regularity lemma and its applications in graph theory. In: Khosrovshahi, G.B., Shokoufandeh, A., Shokrollahi, A. (eds.) Theoretical Aspects of Computer Science: Advanced Lectures, pp. 84{112. Springer, Berlin (2002)

3. Szemeredi, E.: Regular partitions of graphs. In: Colloques Internationaux CNRS 260|Problemes Combinatoires et Theorie des Graphes, pp. 399{401. Orsay (1976)

The Generalized Traveling Salesman Problem

Mariia Gordenko

National Research University Higher School of Economics

The Generalized Traveling Salesman Problem is an extension of Travelling Salesman Problem. The list of cities and their distances is separated into clusters (one cluster consists from one or more cities). The tour contains only one vertex from each cluster and it is closed. In this research, the different approaches for solving GTSP are given. Potentially useful applications where the problem can be used are presented.

Mapping of politically active groups on social networks of Russian regions (on the example of Karachay-Cherkess Republic)

Galina Gradoselskaya, Ilia Karpov, Tamara Shcheglova

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

The report will show which segments constitute social and political activity in social networks in the KChR region, and how widely it is represented. It is also necessary to define key groups and actors in social networks that create the informal information space of the Republic. We examined groups, persons and media in different social networks. We used the author's methodology for mapping social networks at the federal and regional levels of the Russian Federation in the analysis, tested in several projects. The mapping of social networks was carried out using the method of seed clustering. Collection period: April-May 2017. Collection networks: Facebook, Vkontakte, Instagram, Odnoklassniki, LiveJournal. Eight main clusters of political activity in social networks of KChR were obtained by the author's method of seed clustering (the total number of groups is more than 2000, the groups dedicated only to KChR are more than 200): - The Karachay-Cherkess cluster includes socio-political groups dedicated to the KChR, the media, and groups in support of the opposition (as well as the public movements supported by it, such as the RKNK, Elbrusoid, etc.). The cluster connects both with the common Caucasian Islamic cluster, and with the opposition cluster, and the Stavropol clusters. - The Abkhazian is a secular cluster, partly in the national language. It is devoted to the problems of Abkhazia and the living conditions of Abkhazians in other republics of the Caucasus. - Adyghe is mostly nationalistic, half – in the national language. It also contains socio-political groups and a social movement. The main idea is "the unification of divided nations". - Kabardino-Balkar – mostly nationalistic, partly in the national language. It also includes socio-political groups and social movements. - All-Caucasian, Islamic – there are groups of Caucasian republics: Chechnya, Ingushetia, Dagestan. The main content concerns unification based on one religion. Some groups are purely Islamic, propagandizing lifestyle ("Islamic family", "Islamic values", "Islamic medicine", "Harassment of Muslims", etc.). Partially in national languages. - All-federal opposition

– federal groups, met in earlier research projects when mapping opposition groups at the federal level or at Moscow level. They coordinate the activity of the opposition throughout the country (including the KChR). - Stavropol cluster dedicates to the neighboring region – Stavropol Territory. The content has mostly informational character. - Pan-Turkism in the Caucasus – the cluster based on the activity of Turkish information resources. Partly in Turkish, partly in Russian, partly in the national languages of the Republic. Groups from this cluster promote the idea of uniting all the Turks and territorial claims to the Russian territories (Lower Volga Region, Siberia, and the Caucasus). About a third of the groups in this cluster are directed to work in the KChR Republic (it contains the name of the Republic or its peoples in the group name). In the report, each cluster explicitly will be analyzed by network methods, the most influential persons and social movements are shown, the content of their information and network activity is analyzed.

FPT-algorithms for some problems related to integer programming

Dmitry Gribanov

National Research University Higher School of Economics
https://nnov.hse.ru/en/latna/

In this report, we will consider FPT-algorithms for special cases of the shortest lattice vector, closest lattice vector, integer linear programming, and simplex width computation problems, when matrices included in the problems' formulations are near square. The parameter is the maximum absolute value of rank minors of the corresponding matrices. Additionally, we will consider FPT-algorithms with respect to the same parameter for the problems, when the matrices have no singular rank sub-matrices. Finally, we will try to apply previous results for the empty lattice simplicies enumeration.

Advances in the enumeration of regular graphs and not-so-regular graphs

Mikhail Isaev

Monash University, Australia
Moscow Institute of Physics and Technology, Russia

joint work with Brendan D. McKay (Australian National University)

Problems of counting graphs with given degree sequence arise naturally in network analysis where the only data available are collective properties of network nodes. Determining exact values for the counts of interest is usually a difficult task, even for such a fundamental question as the number of regular graphs there is no computationally efficient exact formula available in the literature when the degree $d > 4$. Asymptotic enumeration of regular graphs has been the subject of extensive study since Read solved the case $d = 3$ in his 1958 PhD thesis. Until recently the best result for sublinear d in the number of vertices n was obtained by McKay and Wormald in 1991, who proved an asymptotic formula for $d = o(n^{1/2})$. Even earlier, McKay and Wormald used complex-analytic methods in 1990 to show that the same asymptotic formula holds for $d \geq n/\log n$. The gap between these two domains remained open for more than a quarter-century, until Liebenau and Wormald closed it with an elegant argument in 2017.

The aforementioned results considered not only regular graphs, but graphs having some limited amount of irregularity in the vertex degrees. They are general enough to cover typical degree sequences that appear in the standard random graph models with high probability. In the case where the vertex degrees are linear, an even wider range was achieved by Barvinok and Hartigan in 2013.

In this talk we describe an extension of the complex-analytic approach that works for average degree at least n^c for any $c > 0$ and allows a large amount of variation of degrees as well as specification of a possibly large forbidden subgraph. It is based on a bound of Isaev on the truncation error for the cumulant generating function of a complex random variable. In the case of regular graphs, we show that the method leads to an asymptotic expansion and determine the first few coefficients.

Algorithms of Extracting Patent-Related Data from Professional Blogs to Improve Patent Value Prediction

Alexander Ivanov[1,2], Zeljko Tekic[1]

[1] Skolkovo Institute of Science and Technology,
[2] National Research University Higher School of Economics

With an increase in the relevance of knowledge assets, inventions and innovations for doing business and promoting economic growth, intellectual property (IP) rights, and especially patents, started receiving significant attention. That resulted in an ever-increasing number of patent applications and patents in power - at the moment, there are more than 10 million patents in force, and more than 1.3 million new applications every year. However, the value of patents is not uniform. It is highly skewed to the left – only few patents are valuable, and a majority of them are associated with very little or no value [1][2]. And knowing which patents are valuable is very important in technology and strategic management. Why? Because if we know which patents are valuable, we know which patents will drive development of technology and standard setting in an industry; which patents will help in building temporary monopoly and opportunity for good licensing deals; which patents influence a company value, and when we should do mergers and acquisitions. Furthermore, it can help to determine damages in litigations and in optimization of patent portfolios. Thus, the question of measuring and predicting patent value have been in the focus of economists, business scholars, practitioners and policy makers for some time.

The literature suggests that most frequently used patent value indicators are those that could be readily extracted from publicly available databases, like number of forward and backward citations, patent family size, number of different IPC classes assigned to the patent, patent life-time, and the number of claims [3]. Among these patent value indicators, the number of patent's forward citations had been validated as the most important (reliable and exploratory) indicator of the patent value (e.g.: [2][4][5]). However, this indicator has one major problem – a patent needs time to pass, between several months to several years after publication, to receive citations from other patents. Thus, its usage in predicting value is limited. A recent paper [6] shows

that it may be possible to use the time of the first citation ("first patent citation lag") as proxy to the total number of citations, and as patent value predictor. But, even in this case at least several months are needed for a patent to get its first citation.

In our research we propose a novel approach for predicting patent value early in their life-time. This approach is based on usage of contextual information in patent analytics. Contextual information for us is information that is related to a specific patent, but it is not part of patent (document or database) itself, for example blog and social media posts that discuss specific patent. Following this idea, we try to identify, collect and analyze media entries discussing published patents on the Internet.

In this paper, we discuss our system for identifying and collecting patent-related information published on the Internet. In order to collect blog entries about patents, a software tool has been developed. The crawling algorithm is based on BFS (breadth-first search). It takes the list of websites as input, analyzes each page on the websites and extracts articles about specific patents. Extraction of articles which describe a specific patent is based on the keyword search, regular expressions and hyperlink parsing. The result of the BFS search gives additional external data sources, most popular of which are added to the list of data sources allowing the search to be self-sustainable in finding news websites with information about patents.

How this method can be applied? A tool for the search of the recent valuable patents can be developed. Small and medium companies which do not have large budgets for intellectual property management can use this tool. Instead of looking at thousands of patents published every day, they can get notifications of only the most potentially interesting patents found by our software.

We report on the current results and technical details of the blog search with 400+ websites in the list of sources. We present the search model, the software structure and the algorithms, the challenges which we were facing during the implementation procedure and during the test phase. Details will be given in the presentation at the conference.

References

[1] D. Harhoff, F. Narin, F. M. Scherer, and K. Vopel, "Citation Frequency and the Value of Patented Inventions," *Rev. Econ. Stat.*, vol. 81, no. 3, pp. 511–515, 1999.

[2] B. Gambardella, A, Harhoff, D and Verspagen, “The Value of European Patents,” London, 2008.
[3] Z. Tekic and D. Kukolj, “Threat of Litigation and Patent Value: What Technology Managers Should Know,” *Res. Manag.*, vol. 56, no. 2, pp. 18–25, 2013.
[4] D. Harhoff, F. M. Scherer, and K. Vopel, “Citations , Family Size , Opposition and the Value of Patent Rights,” 2002.
[5] W. J. Allison J., Lemley M., “Extreme Value or Trolls on Top - The Characteristics of the Most-Litigated Patents,” 2009.
[6] J. Lee and S. Y. Sohn, “What makes the first forward citation of a patent occur earlier?,” *Scientometrics*, pp. 1–20, 2017.

Loss function, unbiasedness, and optimality of Gaussian graphical model selection

Valery Kalyagin, Petr Koldanov, Alexander Koldanov, Panos M. Pardalos

National Research University Higher School of Economics
https://nnov.hse.ru/en/latna/
and
University of Florida

A Gaussian graphical model is a graphical representation of the dependence structure for a Gaussian random vector. Gaussian graphical model selection is a statistical problem that identifies the Gaussian graphical model from observations. There are different statistical approaches for Gaussian graphical model identification. However, general statistical properties of resulting statistical procedures, such as unbiasedeness and optimality, are not well known. In this paper we study these properties. We consider the graphical model selection problem in the framework of multiple decision theory and suggest measuring the quality of statistical procedures using an additive loss function. Associated risk function in this case is a linear combination of the expected numbers of two types of errors (False Positive and False Negative). We discuss the general concept of unbiasedeness of multiple decision procedure and its projection on individual tests. To find the optimal statistical procedure, we combine the tests of a Neyman structure for individual hypotheses with simultaneous inference. We show that the obtained multiple decision statistical procedure is optimal in the class of unbiased multiple decision procedures.

Optimization of Network Schedule of Set of Operations in Excel

Nikolai Katargin

Financial University under the Government of Russian Federation

The method of optimization of the network schedule by time and number of employees using the Excel service and the method of building the density of the distribution of the project end time by Monte Carlo with Visual Basic method at any distribution of the probabilities of the work time are proposed.

On equivalence of network structures in different random variables networks.

Petr Koldanov

National Research University Higher School of Economics
https://nnov.hse.ru/en/latna/

Problem statement: network model of comlex system is a complete weighted graph with nodes corresponds to the elements of the system and weights of edges are given by some measure of connection between them. Network models are widely used in the stock market network analysis [1],[2],[3]. In the case different network structures which contain a key information of network models are analyzed. Minimum spanning tree (MST) [1], planar maximally filtered graph (PMFG) and market graph [2], [3] are most popular network structures. Key problem is to identify these network structures by observations of complex system elements.

Traditional approach to the problem is to calculate Pearson correlation for any pair of stocks and to apply corresponding algorithms to network structures identification.

In the presentation the concept of random variables network is introduced. Random variables network is a pair (X,γ), where vector $X=(X_1,...,X_p)$ has multivariate distribution and $\gamma=(\gamma_{ij}=\gamma(X_i,X_j)),i,j=1,...,p$ - is a measure of similarity between X_i,X_j. It is easy to see that traditional approach is based on application of corresponding algorithms of network structures identification to Pearson correlation network. In the presentation the sign random variables network is introduced which is based on probability of sign coincidence of two random variables. It is shown that if vector $X=(X_1,...,X_p)$ has multivariate elliptically contoured distribution then network structures in Pearson correlation network and network structures in sign correlation network are coincide. The procedures for network structures identification in sign correlation network are constructed. It is proved that these procedures have invariant risk function. This work is partially supported by RFFI grant 18-07-00524.

References:

1. Mantegna, R.N. Hierarchical structure in financial markets //The European Physical Journal B-Condensed Matter and Complex Systems. 1999. Vol.11, issue 1. Pp. 193-197.
2. Boginski V., Butenko S., Pardalos P.M. Statistical analysis of financial networks, Computational Statistics and Data Analysis. 48 (2), 431--443 (2005).
3. Boginski V., Butenko S., Pardalos P.M. Mining market data: a network approach J. Computers and Operations Research. 33 (11) 3171--3184 (2006).
4. Kalyagin V.A., Koldanov A.P., Koldanov P.A., Pardalos P.M., Zamaraev V.A. Measures of uncertainty in market network analysis, Physica A: Statistical Mechanics and its Applications, v. 413, No. 1, pp. 59-70 (2014).
5. Koldanov, P. M. Pardalos, V. A. Zamaraev Statistical uncertainty of market network structures // DATA ANALYTICS 2014, The Third International Conference on Data Analytics. -2014, pp. 91-94.
6. Koldanov P.A. Risk function of statistical procedures for network structures identification // Vestnik TvGU. Seriya: Prikladnaya Matematika [Herald of Tver State University. Series: Applied Mathematics], 2017, no. 3, pp. 45–59. (in Russian). https://doi.org/10.26456/vtpmk178

Continuous connectome model of brain network exhibits scale-free property.

Anvar Kurmukov

National Research University Higher School of Economics

In the field of neuroscience, networks naturally arise in the form of brain connectivity map. While it is not possible to obtain edges at a fiber (neuronal) scale, modern connectomics considers the level of anatomically segregated brain regions to be connected by inter-regional pathways. The delineation of brain regions (or parcellation) is critical. Some studies consider a set of homogeneously distributed and equal-sized regions, while others define parcels based on anatomical structure. All in all, there are dozens of different brain parcellations. Networks built upon these parcellations have a different number of vertices, degree distribution, and other properties, but all represent the same brain connectivity. In recent works a new concept of a continuous connectome was introduced (D. Moyer et al. 2017, B. Gutman et al. 2014). It is a general framework of cortical brain connectivity where any particular parcellation of the cortex may be expressed by the derived connectivity matrix of continuous connectome model. Here we will demonstrate some interesting properties of a continuous connectivity model, particularly a power-law node distribution, which is generally not typical for structural connectomes.

Individual project effectiveness: individual network predictors

Valentina Kuskova, Ivan Kuznetsov, Vladimir Kuznetsov, Rustam Kamalov

National Research University Higher School of Economics

In organizational studies, the role of social networks has occupied the place of prominence for quite some time. Both the theories of social selection and social influence have been tested for their explanatory power of individual and organizational effectiveness. However, to the best of our knowledge, little attention has been paid to the individual project effectiveness from the network perspective. This is largely due to the fact that such studies need to be longitudinal, placing a lot of demand on data collection efforts. In this study we were able to overcome the methodological limitations by using longitudinal design with three waves of data collection in order to test the models of network influence on project effectiveness outcome over a period of a year. Our results offer both the individual and the organizational practical implications and examine the role of individual networks from the theoretical perspective.

Co-authorship recommender System and Graph Embeddings

Ilya Makarov

National Research University Higher School of Economics

We consider new formulation of graph embedding algorithm and its applications to formulating co-author recommender system as a link prediction problem on attributed co-authorship network. We evaluate our approach on the co-authorship network of HSE researchers publications improving previous results using manual feature engineering. We compare it with existing structural network embeddings and feature-engineering models.

Network structure and scheme analysis of the Russian-language segment of Wikipedia

Sergey Makrushin

Financial University under the Government of the Russian Federation

Nowadays Wikipedia is much more than the most popular online encyclopedia: it is a unique source of semi-structured knowledge about the world. Data from Wikipedia is actively used in different approaches to create universal semantic networks. In the research we created the network of the Russian-language segment of Wikipedia and made analysis of its structure. In our network Wikipedia articles and categories are regarded as nodes and references as links. The emphasis of the research is on studying nodes degree distribution. In the work it is suggested to use adaptive kernel density estimation method for clearing diffusion of tails of nodes degree distributions. Using this method, we found that nodes degree distribution for articles out- links doesn't fit to power law and has extensive artifacts. We found that these artifacts were induced by using navigation templates in Wikipedia articles. For eliminating these artifacts, we suggest to include into the Wikipedia network scheme new type of nodes and links representing Wikipedia templates and their references.

Indirect influence assessment in the context of retail food network

Natalia Meshcheryakova

National Research University Higher School of Economics,
Institute of Control Sciences of RAS

In current work we propose some algorithms of long-range connections evaluation. Firstly, we reconstruct an initial graph into the graph of directed intensities. This approach is based on the individual and the group influence with respect to nodes characteristics. Secondly, we apply different models of the indirect influence estimation based on simple paths and random walks. This can help us to estimate node-to-node influence in networks. Finally, we aggregated the results of node-to-node influence into the influence index. The model is applied to the food trade network based on World International Trade Solution database. The results are compared with classical centrality measures that were applied to initial network and to the network of directed intensities.

Metaheuristcs for Large-Scale Graph Clustering

Pierre Miasnikof[1], Leonidas Pitsoulis[2], Anthony J. Bonner[3] and Yuri Lawryshyn[1]

[1] University of Toronto, Dept. of Chemical Engineering and Applied Chemistry, Toronto, Canada
[2] Aristotle University of Thessaloniki, Dept. of Electrical and Computer Engineering
[3] University of Toronto, Dept. of Computer Science, Toronto, Canada

Graph clustering is an NP-hard problem. In the context of ever increasing volumes of data, scalable clustering techniques remain an open and actively researched problem. In this article, we formulate the clustering problem as a mathematical optimization exercise. Our objective function is inspired by the "similarity measure" introduced by Fan and Pardalos [2], but does not require cardinality constraints for each individual cluster and is more numerically stable than the commonly used
modularity (e.g., [1]).

After modifying to the objective function that allows us to eliminate individual cluster cardinality constraints, we explore its application to very large graphs. We attempt to solve our formulation using three commonly used metaheuristic techniques, a simple greedy search algorithm, GRASP (greedy algorithm combined with local search) and simulated annealing. We compare their performance, paying particular attention to scalability. We use simulated data with known optima as benchmarks.

References

1. Aloise, D., Caporossi, G., Hansen, P., Liberti, L., Perron, S., Ruiz, M.: Modularity maximization in networks by variable neighborhood search. In: Bader, D.A., Meyer-
henke, H., Sanders, P., Wagner, D. (eds.) Graph Partitioning and Graph Clustering, 10th DIMACS Implementation Challenge Workshop, Georgia Institute of Technology, Atlanta, GA, USA, February 13-14, 2012. Proceedings. pp. 113–128 (2012),
http://www.ams.org/books/conm/588/11705

2. Fan, N., Pardalos, P.M.: Linear and quadratic programming approaches for the general graph partitioning problem. J. of Global Optimization 48(1), 57–71 (Sep 2010), http://dx.doi.org/10.1007/s10898-009-9520-1

Social Mechanisms of the Subject Area Formation. The Case of “Digital Economy”

Alexander Kharlamov[1,2], Galina Gradoselskaya[2], and **Oxana Mikhailova**[2]

[1] Institute of Higher Nervous Activity and Neurophysiology of RAS, Moscow, RF
Moscow State Linguistic University, Moscow, RF

[2] National Research University Higher School of Economics, Moscow, RF

The structure of natural language could be considered as a social network. In the complete network we can allocate speech markers, which describe the subject and semantic areas. In this article a wide range of texts about digital economy is analyzed making it possible to show the thematic structure of this subject area. Central and peripheral concepts were identified to characterize theoretical core concept and related topics clarifying the application of digital economy. Identification of thematic areas is performed in two ways – through the construction of a thematic tree through neural network modeling in the Text Analyst and through the analysis of semantic networks. The results, approaches and methods of this study can be used in considering other large thematic fields related to new ideological currents being developed as an element of social design and management.

Networks: History, The Present, and Future Challenges.

Panos M.Pardalos

University of Florida
www.ise.ufl.edu/pardalos
and
National Research University Higher School of Economics
https://nnov.hse.ru/en/latna/

The network revolution began in the 18th century when Euler solved the famous Konigsberg bridge problem. In the 19th century, Kirchhoff initiated the theory of electrical networks and was the first person who defined the flow conservation equations, one of the milestones of network flow theory. After the invention of the telephone by Alexander Graham Bell in the 19th century, the resulting applications stimulated network analysis.

The field evolved dramatically after the 19th century. At present, our lives are both affected by and interface with the networks that connect us.

After a brief historical overview, the talk will focus on recent exciting developments and discuss future challenges not only in the science of networks but also in our network-driven and changing society.

Grouping Games: Finding Clusters in Graphs, Digraphs and Hypergraphs

Marcello Pelillo
Ca' Foscari University of Venice, Italy

Clustering refers to the process of extracting maximally coherent groups from a set of objects using pairwise, or high-order, similarities. Traditional approaches to this problem are based on the idea of partitioning the input data into a predetermined number of classes, thereby obtaining the clusters as a by-product of the partitioning process. In this talk, I'll provide a brief review of recent work done in my group which offers a radically different view of the problem. In contrast to the classical approach, in fact, we attempt to provide a meaningful formalization of the notion of a cluster and we show that game theory offers an attractive and unexplored perspective that serves well our purpose. To this end, we formulate the clustering problem in terms of a noncooperative "clustering game" and show that a natural notion of a cluster turns out to be equivalent to a classical (evolutionary) game theoretic equilibrium concept. We prove that the problem of finding the equilibria of our clustering game is equivalent to locally optimizing a polynomial function over the standard simplex, and we provide a discrete-time dynamic to perform this optimization, based on the Baum-Eagon inequality. The proposed grouping framework, which has already found applications in a variety of application fields, including computer vision, bioinformatics, security and video surveillance, etc., is general and can be applied to weighted graphs, digraphs and hypergraphs alike.

A Method for Overlapping Community Detection in Networks based on Link Partitioning and Partitioning Around Medoids

Alexander Ponomarenko

National Research University Higher School of Economics
https://nnov.hse.ru/en/latna/

In this talk a new method will be presented for detecting overlapping communities in the networks with predefined number of clusters. The method is a combination of link partitioning and partitioning around medoids as a method for detecting disjoint communities on the linear graph. In turn, partitioning around medoids is done based on one of the distance function defined on the set of nodes of the linear graph. In this talk the generalize degree distance, commute distance and shortest path distance function are considered as distance function. Strong and weak points of the proposed method will be discussed through the study of real examples.

On maximum degree-based quasi-clique problem

Oleg A. Prokopyev

University of Pittsburgh
and
National Research University Higher School of Economics
https://nnov.hse.ru/en/latna/

We consider the problem of finding a degree-based γ-quasi-clique of maximum cardinality in a given graph for some $\gamma \in (0; 1]$. A degree-based γ-quasi-clique (often referred to as simply a quasi-clique) is a subgraph, where the degree of each vertex is at least γ times the maximum possible degree of a vertex in the subgraph. A degree-based γ-quasi-clique is a relative clique relaxation model, where the case of $\gamma = 1$ corresponds to the well-known concept of a clique. In this talk we first discuss the computational complexity of this problem and then describe exact methods for its solution.

Quantification of graph dissimilarities using Information Theory

Martín Gómez Ravetti

Federal University of Minas Gerais, Brazil

In this talk, we review a novel method to quantify graph dissimilarities. The pseudo-metric is an efficient and precise measure for network comparison, based on the quantification of differences between the distance probability distributions extracted from them. The measure can identify and quantify structural topological differences that have a practical impact on the information flow through the network, such as the presence or absence of critical. An extension to labeled graphs is also presented. In the case of multiplex networks, each layer possesses its own connectivity configuration, and the quantification of their dissimilarities is associated with the concept of “diversity". Layers with more dissimilar connectivity patterns create a more diverse multiplex structure. Several applications on real and artificial data are examined, exemplifying the usefulness of the measure in a variety of systems.

Dichotomy Complexity of Graphs: Notions and Algorithms

Alexander Rubchinsky

National Research University Higher School of Economics,
National University of Science and Technology MISIS, Russian Federation

The report is devoted to the relatively new notion of dichotomy complexity of graphs. Its origin is connected to the following algorithm 1 of graph dichotomy:

1. Set the current frequency at every edge equal to 1.
2. Randomly choose two different vertices of the graph.
3. Find a path between the two vertices. If such a path does not exist, go to step 7.
4. Add one to frequencies in all the edges on the path found at step 3.
5. If the number of consecutive runs of steps 2 – 4 does not exceed a pre-specified number *T*, go to step 2. Otherwise, go to the next step 6.
6. Remove an edge with the maximum frequency and return to step 1.
7. Graph *G* is divided into two components, which form the found decomposition.

The described algorithm differs from the well known Newman-Girvan algorithm [1] only in one “technical” aspect: instead of all the pairs of vertices of a given graph, only randomly chosen pairs of vertices have been considered. In many cases the results of dichotomy remains the same for both op-tions. However, in more complicated (and more real) cases the modified algorithm has led to different dichotomies of the same graph. Firstly these cases have been considered as vexatious mistakes. But it is not. Moreover, it is possible to assert that numerous dichotomies naturally arise in the study of many graphs. Because many real systems are naturally modeled by graphs, the domain of applications of graph dichotomies is large enough.

A family of dichotomies can be used in analysis of various complex systems as follows. Assume the family consists of *M* items. Some of them can coincide and some of them can be different. Assume that among *M* found dichotomies dichotomy d_p is encountered m_p times (p = 1, …, t), where $\sum_{p=1}^{t} m_p = M$. Numbers m_p (p = 1, …, t) are calculated directly from the family of *M* dichotomies. Let us define

$$E = -\sum_{p=1}^{t} \mu_p \ln\left(\mu_p\right), \text{ where } \mu_p = m_p / M.$$

E is the conventional entropy of division of M items into t parts, consisting of coinciding items. It is obvious that the minimal possible value of E is 0, while the maximal possible value is $\ln(M)$. Finally, assume $I = E / \ln(M)$. Thus, I is a random value, expressing complexity, entanglement, intricacy and other similar hardly defined, though important, properties of various real systems. Minimal possible value 0 corresponds to completely ordered systems, while maximal value 1 corresponds to completely chaotic ones. Therefore, construction of the above mentioned family of dichotomies is an important problem.

Shortcomings of algorithm 1 are obvious. Any return to step 1 implies loss of all the accumulated data about edge frequencies. Hence, the algorithm req uires too many operations in order to find even one dichotomy, and, therefore, it cannot be used for real dimensions.

In algorithm 1, paths, connecting a next pair of vertices, are traced ***independently of all the already traced paths***. Yet, taking into account all the already traced paths can obtain a cut between two sets of vertices, who's all the edges have the ***same maximal frequency***. Then concurrent removal of all the edges with the maximal frequency defines the desired dichotomy of the graph. The corresponding algorithm 2 can be shortly described as follows.

1. *Initial setting*. Initialize frequencies in all the edges by 1.

2. *Cumulative stage.* All the operations of this stage are repeated T times:

2.1. Random choice of a pair of different vertices of graph G.

2.2. Construction of a minimal path (connecting the two chosen vertices so that its longest edge is the shortest one among all such paths) by Dijkstra algorithm

2.3. The value 1 is added to the frequencies in all the edges on the path found at the previous step.

3. *Final stage*.

3.1. The maximal value of frequency $f_{\max}$ over all the edges is stored.

3.2. The operations of steps 2.1 – 2.3 from cumulative stage are executed till the current maximal value does not exceed $f_{\max}$.

3.3. Deduct the value 1 from the frequencies in all the edges forming the last found path.

3.4. Remove all the edges, in which frequency is equal to $f_{\max}$.

3.5. Components of the modified graph form the required dichotomy.

Algorithm 2 is much more efficient than algorithm 1. But execution of algorithm 2 under any number of steps T at cumulative stage produces only one dichotomy. However, in order to achieve a reasonable accuracy in a

graph with 450 vertices we must assume T is equal to 10000 and M (the number of found dichotomies) is equal to 1000. Therefore we encounter the same problem: how to essentially contract the time of construction of a reasonably large set of all the possible dichotomies? The answer is relatively simple. During the execution of cumulative stage of algorithm 2 there are many times when the maximal value of frequency f_{max} over all the edges is increased. At every such a case we can execute operations 3.3 – 3.5 of the final stage, and, hence, construct a dichotomy.

The algorithm 3 executes some a priory unknown number of operations 2.1 – 2.3 till the number of constructed dichotomies becomes equal to a pre-specified number (say, 1000). The experiments demonstrate that the time is contracted significantly – in hundreds times.

The results concerning 1st and 2nd algorithms are presented in [2] and in the cited there publications. The 3rd algorithm is the new one. As before, despite the technical character of changes the new algorithm allows stating and solving problems that could not be considered before.

Citations

1. Girvan, M., Newman, M.E.J. Community structure in social and biological networks. – Proc. Natl. Acad. Sci. USA 99, 7821–7826, 2002.

2. Rubchinsky A.: A New Approach to Network Decomposition Problems. In “Models, Algorithms, and Technologies for Network Analysis”. Springer Proceedings in Mathematics & Statistics. 2017, pp. 127 – 152.

Differentiation of discursive political space in social networks

Galina Gradoselskaya, Ilia Karpov, **Tamara Shcheglova**

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

The paper analyzes speech markers and semantic concepts typical for patriotic and oppositional discourse in social networks. It is necessary to emphasize that the terms "oppositionists" and "patriots" in our study are conventionally accepted, by the principle of self-identification by representatives of these groups, and by their labeling of groups of opponents. The aim of the study is to determine the speech markers and concepts peculiar for the identified political groups - the bearers of certain political attitudes. 230 patriotic and 240 oppositional resources were expertly selected in three social networks: Facebook, VKontakte and LiveJournal. About 100,000 posts from these resources were analyzed and 35,000 most frequent speech markers were processed, of which 1800 markers were selected for analysis. The alternative method to tf-idf metric for specific text markers identification is proposed. The features of oppositional discourse in comparison with the patriotic discourse were formulated. The results of the research can have both scientific-methodological and applied significance. On the one hand, the analysis of sets of speech markers that characterize political groups allows us to understand social models and attitudes embedded in the discourse and the subsequent behavior of representatives of these groups. On the other hand, it is possible to extend a set of keywords for text search of a certain political orientation, based on the obtained results. In addition, 600 largest politically active groups in Vkontakte were selected. On the basis of joint participation in the groups, a connection graph was drawn and key clusters were identified. Oppositional cluster and its discourse placed in the structural center, dictating fashion to other political currents.

Computational Complexity of SRIC and LRIC indices

Sergey Shvydun

National Research University Higher School of Economics, ICS RAS

Over the last years, there is a deep interest in the analysis of different communities and complex networks. The problem of identification of key elements and the most important connections in such networks is one of the main areas of research. However, the growth of sizes of real networks as well as their heterogeneity makes the problem both important and problematic. The application of short-range and long-range interaction centrality (SRIC and LRIC correspondingly) indices can be used to solve the problem since they take into account different important aspects such as individual properties of nodes, the possibility of their group influence and topological structure of the whole network. However, the computational complexity of such indices needs a further consideration. Thus, our main focus is on the performance of SRIC and LRIC indices. As a result, we proposed several modes how to decrease their computational complexity of the indices. The runtime comparison of the sequential and parallel computation of the proposed models is also given.

Recent applications of Variable Neighborhood Search

Angelo Sifaleras

Department of Applied Informatics, School of Information Sciences,
University of Macedonia, Thessaloniki, Greece

joint work with I. Konstantaras

Over the past two decades, reverse logistics and closed-loop supply chain networks have gained substantial interest in business and academia. The dynamic lot sizing problem with product returns and recovery is one of the most extensively researched topics in inventory control literature. Several interesting generalizations of this optimization problem have lately emerged that include the multi-product case, the case with capacity constraints, and others. In this lecture, we present recent successful applications of Variable Neighborhood Search for the efficient solution of such problems, review the state-of-the-art solution methods, and also discuss some open research problems.

Related references:

- Sifaleras, A., & Konstantaras, I. (2017). Variable neighborhood descent heuristic for solving reverse logistics multi-item dynamic lot-sizing problems. Computers & Operations Research, 78, 385-392.
- Sifaleras, A., Konstantaras, I., & Mladenović, N. (2015). Variable neighborhood search for the economic lot sizing problem with product returns and recovery. International Journal of Production Economics, 160, 133-143.

On local graph transformations for the vertex k-coloring problem

Dmitry Sirotkin

National Research University Higher School of Economics
https://nnov.hse.ru/en/latna/

Graph transformations are widely used to determine whether certain graph problem is solvable in polynomial time for different graph classes. Also they can be helpful for the task of developing polynomial-time algorithms for said problems. The trivial example for such transformation is the replacement of any edge in the graph with the path of length of three, while solving the independent set problem. This adds one to the size of maximum independent set for the graph. We investigate such transformations for the vertex k-coloring problem. Let G be a graph with induced subgraph H. We call subset $A \subseteq V(H)$ H-separator if no vertex from $H \setminus A$ is adjacent to a vertex from $G \setminus V(H)$. Let G be a graph with induced subgraph G_1, which has G_1-separator A and let G_2 be a graph for which $A \subseteq V(G_2)$. We call formation of graph with set of vertices $(V(G) \setminus V(G_1)) \cup V(G_2)$ and set of edges $(E(G) \setminus E(G_1)) \cup E(G_2)$ a replacement of graph G_1 with graph G_2.

In the speech we report on some results which define the nature of such replacements for the k-coloring problem. For any given induced subgraph with corresponding separator we provide a function which attributes this subgraph into corresponding equivalence class. For each of such classes we construct a graph from this class. This means, that for any subgraph in any given graph we can replace it with pre-defined subgraph such that the solution of k-coloring problem does not change. Also we determine the upper bound for the number of vertices in such replacement.

Max-Min Fairness and its Applications

Theodore Trafalis

School of Industrial and Systems Engineering, University of Oklahoma, Norman, OK 73019, USA

In this talk I present fair optimization models and methods applied to networks. In particular I discuss the lexicographic max-min optimization models known as max-min fairness (MMF) models. I also present applications of fair optimization methods for resource allocation problems and communication networks. In addition the problem of controlling congestion on complex networks with respect to fairness, efficiency and network structure is discussed.

Integer programming techniques for modeling clique relaxations in networks

Alexander Veremyev
University of Central Florida

The talk is based on *joint work with* Vladimir Boginski, Sergiy Butenko, Oleg Prokopyev and others.

Clustering is one of the most important and challenging problems in network analysis arising in many practical applications. It has led to the development of various graph-theoretic concepts of clusters based on so-called clique relaxations (subgraphs with a 'relaxed' clique defining property), which in many cases can be naturally described using linear integer programming models. As modern optimization software packages (Gurobi, Fico Xpress, CPLEX, etc.) have been demonstrating significant performance enhancements over the last decade, these improvements allow finding solutions for many clustering problems on real-life and synthetic networks within a reasonable time via developing efficient mixed integer programming (MIP) formulations. In this talk we will overview various clique relaxations and MIP techniques for their modeling and optimization that we have recently developed. Many potential applications, effectiveness and flexibility of the proposed MIP formulations as well as the importance of analytical bounds will also be discussed.

Advisors Role in Russian Education Policy: Application of Network Analysis

Dmitry Zaytsev

National Research University Higher School of Economics

According to the well-known policy advisory system science theory (Halligan 1995, Howlett 2013) the type of brokerage is one of the key factors that determine the impact of policy advisors on policy change. Using the network approach, we will test the position that policy advisors in Russian education policy (as in the case of introduction of the Unified State Exam) occupy between the government and the public. Previous work in the field has looked at the role that various stakeholders could have played in changing educational policies (e.g., Pinheiro, 2015), yet actual brokerage relationships were not explicitly hypothesized. In this study we argue that policy advisors play the broker role between the government and the society, becoming the agents of change on the governmental side first, and then convincing the society of the necessity of new policy decisions. Using the relational approach afforded by network analysis, we examine the types of centrality for various policy advisors in an effort to determine their role in educational policy reforms.