RUSSIAN WORKSHOP ON COMPLEXITY AND MODEL THEORY

Abstracts

June 9–11, 2019 MIPT Moscow, Russia
Workshop on graphs, networks and their applications
ISBN 978-5-6041-1875-7
Organizing Committee
Andrei M. Raigorodskii (MIPT, Yandex)
Vladimir Remeslennikov (Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences)
Maksim Zhukovskii (MIPT)

MIPT

Algorithms

Moscow Institute of Physics and Technology

 Algorithms an Open Access Journal by MDPI

Yandex

Huawei
## CONTENTS

### PLENARY TALKS

**Erich Grädel** Provenance Analysis for Logic and Games ................................................................. 8

**Bruno Courcelle** Structured graphs and the verification of their monadic second-order properties by means of automata ................................................................. 9

**Vladimir Remeslennikov** Switch transformations and a triangle-creation process in regular graphs ........................................................................................................... 11

**Vitaly Roman'kov** Algebraic cryptology: methods of cryptanalysis via (non)linear decomposition and new protection against them ......................................................... 12

**Anuj Dawar** Approximations of Graph Isomorphism ........................................................................ 14

**Jaroslav Nešetřil** Limits of finite structures ........................................................................................ 15

**Johann A. Makowsky** A Logician’s View of Graph Polynomials ......................................................... 16

**Nikolay Vereshchagin** A Conditional Information Inequality and its Combinatorial Applications ............................................................................................................. 17

**Lance Fortnow** The P vs NP Problem in the Era of Big Data and Fast Computing .............................. 18

**Andrei Bulatov** The complexity of the constraint satisfaction problem ............................................... 19

**Igor Pak** Counting integer points in polytopes .................................................................................. 21

**Noam Nisan** The Communication Complexity of Cake-Cutting ....................................................... 22

**Alexei Miasnikov** Hard instances, Challenger-Solver complexity, and Dehn monsters ................... 23

**Moshe Vardi** The Automated-Reasoning Revolution: From Theory to Practice and Back ............... 24

### TALKS

**Alexander Treyer** Universal equivalence of nilpotent graph groups .................................................. 26

**Mikhail Makarov** Logical complexity of induced subgraph isomorphism for certain graph families .............................................................................................................. 27

**Yury Yarovikov** There are no FO sentences with quantifier depth 4 and an infinite spectrum ............ 28

**Alexey Nikitin** Decidability problems for universal and existential theories of the class of partially ordered sets ................................................................................................. 29

**Artem Shevliyak** On outlier detection with dendrograms. Algebraic approach .................................. 30

**Albert Garetta** Equations in solvable groups ...................................................................................... 32

**Il'nr Khuziev** Distributed search of an antipodal vertex in symmetric Cayley graph over boolean cube .................................................................................................................. 33

**Dmitry Zhuk** On the Complexity of the Quantified Constraint Satisfaction Problem ....................... 34

**Jakub Bulín** Algebraic Approach to Promise Constraint Satisfaction .............................................. 35

**Armin Weiss** The isomorphism problem for finite extensions of free groups is in PSPACE ................. 36

**Evelina Danieyarova** Algebraic geometry over abelian groups ......................................................... 37

**Nikolai Polyakov** Dichotomy theorem in computational social choice theory .................................. 38

**Alexander Zaprjagaev** Linear Orderings Interpreted in Presburger Arithmetic .................................. 40

**Ivan Prikhodko** Communication complexity in T-cell receptors signalling during antigen recognition .................................................................................................................. 42

**Markov Pavel** Analysis of methods for solving of systems of equations for modeling of one- and two-phase flow in network models of pores and capillaries ..................... 43
PLENARY TALKS
A model checking computation checks whether a given logical sentence is true in a given finite structure. Provenance analysis, based on interpretations in commutative semirings, abstracts from such a computation mathematical information on how the result depends on the atomic data that describe the structure. In this approach, atomic facts are interpreted not just by true or false, but by values in an appropriate semiring, where 0 is the value of false statements, whereas any element $a \neq 0$ of the semiring stands for some shade of truth. These values are then propagated from the atomic facts to arbitrary statements in the language. In database theory, provenance analysis has been successfully developed for positive query languages such as unions of conjunctive queries, positive relational algebra, or datalog. However, it did not really offer an adequate treatment of negation or missing information.

We propose a new approach for the provenance analysis of logics with negation, such as first-order logic and fixed-point logics. It is closely related to a provenance analysis of the associated model-checking games, and based on new provenance semirings of polynomials and formal power series, which take negation into account. They are obtained by taking quotients of traditional provenance semirings by congruences generated by products of positive and negative provenance tokens; they are called semirings of dual-indeterminate polynomials or dual-indeterminate power series.

Beyond the use for model-checking problems in logics, provenance analysis of games is of independent interest. Provenance values in games provide detailed information about the number and properties of the strategies of the players, far beyond the question whether or not a player has a winning strategy from a given position.

This is joint work with Val Tannen.
Algorithmic meta-theorems relate the logical expression of a graph property to the complexity of checking it for graphs belonging to certain classes of graphs. Such results actually extend to relational structures. There are many results for first-order (FO) logic and classes of sparse, unstructured graphs, like planar graphs or graphs of bounded degree. However, FO logic is rather weak for expressing graph properties.

Monadic second-order (MSO) logic is much more expressive. This language uses quantifications on sets of vertices and, in some cases, on sets of edges. The graph classes that allow linear-time algorithms for verifying MSO properties are (necessarily) characterized by bounds on tree-width or clique-width. These bounds are parameters in the sense of Fixed Parameter Tractability. They are associated with hierarchical descriptions (i.e. tree-structurings) of the input graphs (of two different, but related types). Linear-time algorithms are obtained for graphs given with the relevant decompositions, that are not easy to compute.

The standard proofs of the meta-theorems construct from MSO sentences and given bounds on tree-width or clique-width, finite automata intended to run on tree-decompositions (expressed by algebraic terms) or on clique-width terms. However, these automata have so many states that their constructions are not usable in practice. This is actually unavoidable (Frick and Grohe, 2004). To overcome this difficulty, we (Irène Durand and myself) have introduced fly-automata. These automata do not store states and transitions but compute them, whenever needed. Only the necessary transitions are computed. They have been implemented and tested, in particular for NP-complete problems about coloring and Hamiltonicity.
The lecture will review: 1) the main results for FO logic, 2) the (well-known) notion of tree-decomposition, 3) the notion of clique-width, a graph complexity measure based on three elementary graph operations, from which graphs are constructed, 4) graph classes for which $\text{tree-width}(G) = O(\text{clique-width}(G))$, 5) the construction of fly-automata for clique-width terms (and the verification of MSO graph properties), and 6) their extensions to counting and optimization problems: for example, computing the number of $k$-colorings, or of $k$-acyclic colorings, of a given graph.

An on-line software (called TRAG, by I. Durand and M. Raskin) implements some predefined fly-automata and the automatic construction of a fly-automaton from an MSO sentence. It also includes the construction of the necessary decompositions for « small » random graphs and for « large » regular ones. A demonstration can be made outside of the main lecture. This is joint work with Irène Durand (LaBRI, Bordeaux University).

References


The main purpose of this report is to give an outline of important results on the universal theory of simple graphs in the formal language of two predicates: the “predicate of equality” $x = y$, and the “neighbourhood predicate” $E(x, y)$, defined on the vertices of the graph $G$.

Summary of the content of the report:

Three equivalences on the vertices of the graph $X = V(T)$ are made, and the concept of a squeezed graph introduced.

The definition of the category of simple graphs with labels is given.

The definition and properties of the three main universal graph invariants are given.

The main universal classes of finite squeezed graphs are described.

Pseudo-Boolean graphs are introduced and the Rado theorem for direct products of pseudo-Boolean graphs is stated.

Universal classes of graphs of finite height and width are defined, and the Rado theorem for them is stated.

As this material is difficult to cover in a single report, associated results, obtained by the application of this approach, will be presented in the reports of E. Daniarova and A. Treyer.
The talk is devoted to algebraic cryptology, that is an area of investigation which operates with algebraic structures (groups, rings and so on) in construction of cryptographic schemes. This area is a part of post-quantum cryptography which develops schemes resistant against attacks via quantum computers of potentially high power.

See [1] and [2] for basic notions and results in algebraic cryptography.

In the first part of the talk, we discuss an attack, termed a linear decomposition attack, on many of known in literature group-based cryptosystems. This attack gives a polynomial time deterministic algorithm that recovers the secret shared key from the public data in the schemes under consideration. Furthermore, we show that in this case, contrary to the common opinion, the typical computational security assumptions are not very relevant to the security of the schemes, i.e., one can break the schemes without solving the algorithmic problems on which the assumptions are based. In a number of monographs and papers, we have shown that in many systems of algebraic cryptography, where the platform group $G$ is a subset in a linear space over a finite or infinite field, we can efficiently solve the computational Diffie-Hellman-like problems and hence to compromise the corresponding cryptographic systems. Other and in some points similar approach was established by Tsaban et al. Also we discuss a non-linear decomposition attack that can be applied in many other cases, in particular, when the platform is a polycyclic group. We present two general schemes for which many schemes are specific realizations. One of these two schemes joins schemes based on two-side multiplications, the second scheme joins schemes based on automorphisms. The two mentioned above attacks show vulnerability of these two general schemes. The main ideas and applications of the (non)linear decomposition methods are presented in papers [3-8] and monographs [9-16] (see also the bibliography in [10]).

In the second part of the talk, we introduce a novel method that is resistant against linear algebra attacks. In particular, we propose an improved version of the famous Anshel-Anshel-Goldfeld algebraic cryptographic key-exchange scheme, that is in particular resistant against the Tsaban et al. linear span cryptanalysis. Unlike the original version, that based on the intractability of the simultaneous conjugacy search problem for the platform group, the proposed version is based on much more hard simultaneous membership-conjugacy search problem and needs to solve the membership problem for a subset of the platform group that can be easily and efficiently built as very complicated and without any good structure. A number of other hard problems should be previously solved by any intruder to start solving of the simultaneous membership-conjugacy search problem to obtain the exchanged key. We also show how this new approach
can be used to improve many schemes based on the conjugacy search algorithmic problem. These results are presented in [11]

REFERENCES.
The problem of deciding whether two given graphs are isomorphic is a well-studied problem in computational complexity as it is one of the few natural problems which is not known to be solvable in polynomial time nor known to be NP-hard. In this talk, I will discuss equivalence relations that over-approximate graph isomorphism but are decidable in polynomial-time. One such class of relations are the Weisfeiler-Lehman equivalences, which have a large variety of equivalent characterisations emerging separately from algebra, combinatorics and logic. I will then discuss equivalence relations that strengthen this by considering invertible maps over finite fields, and present some recent results on them.
Starting from graph limit theory emerged structural limits on the crossroad of model theory, combinatorics and algorithms. We survey recent activity related to structural limits which is both analytic and model theoretic and leads to a surprising connection to new techniques in clustering and modeling of sparsity and stability.

This is a joint work with Patrice Ossona de Mendez (Paris and Prague).
We give a survey on graph polynomials, their expressive power and their complexity of evaluation, based on the following recent papers. We also discuss some conjectures and open problems.


We show that the inequality $H(A | B, X) + H(A | B, Y) \leq H(A | B)$ for jointly distributed random variables $A, B, X, Y$, which does not hold in general case, holds under some natural condition on the support of the probability distribution of $A, B, X, Y$. This result generalizes a version of the conditional Ingleton inequality: if for some distribution $I(X : Y | A) = H(A | X, Y) = 0$, then $I(A : B) \leq I(A : B | X) + I(A : B | Y) + I(X : Y)$.

We present two applications of our result. The first one is the following easy-to-formulate theorem on edge colorings of bipartite graphs: assume that the edges of a bipartite graph are colored in $K$ colors so that each two edges sharing a vertex have different colors and for each pair (left vertex $x$, right vertex $y$) there is at most one color $a$ such both $x$ and $y$ are incident to edges with color $a$; assume further that the degree of each left vertex is at least $L$ and the degree of each right vertex is at least $R$. Then $K \geq LR$. The second application is a new method to prove lower bounds for biclique cover of bipartite graphs.

Joint works Tarik Kaced, Andrei Romashchenko and Nikolay Vereshchagin

*While working on the paper the authors were in part supported by RFBR grants 14-01-93107, 16-01-00362, by an ANR-15-CE40-0016-01 grant RaCAF, and by the Russian Academic Excellence Project ‘5-100’.
The P v NP problem asks if every problem where we can efficiently verify a solution, we can also efficiently find that solution. The P v NP problem is a great mathematical challenge, one of the Clay Math Millennial problems, and we'll review the (limited) progress towards separating P and NP.

The importance of the P v NP came from real-world algorithmic challenges and advances in hardware and algorithms have let us solve NP problems in ways we never thought we could solve years ago. We discuss this progress and the implications across computer science and society in general.
It has been observed long time ago that ‘natural’ computational problems tend to be complete in ‘natural’ complexity classes such as NL, P, NP, or PSPACE. Although Ladner in 1975 proved that if P ≠ NP then there are infinitely many complexity classes between them, all the examples of such intermediate problems are based on diagonalization constructions and are very artificial. Since the seminal work by Feder and Vardi [8] this phenomenon is known as complexity dichotomy (for P and NP), see also Valiant’s work [14] in the context of counting problems. Concerted efforts have been made to make this observation more precise, and since the concept of a ‘natural’ problem is somewhat ambiguous, a possible research direction is to pursue dichotomy results for wide classes of problems. The Constraint Satisfaction problem (CSP) is one of such classes.

Feder and Vardi in [8] observed that the CSP can be conveniently represented as the problem of deciding the existence of a homomorphism between relational structures. In particular, they emphasized the importance of nonuniform CSPs, that is, ones of the form CSP(H), which asks, given a structure G, whether there exists a homomorphism from a structure G to a fixed structure H. This problem belongs to NP in general, but can be solved in polynomial time for some structures H. Early dichotomy results on nonuniform CSPs can be traced back to Schaefer [13] for the Generalized Satisfiability problem (H is 2-element), and Hell and Nesetril [9] for the H-Colouring problem (H is a graph). A systematic study of the complexity of nonuniform CSPs was initiated in [8], where among other advances Feder and Vardi posed the CSP Dichotomy Conjecture, which is the focus of this tutorial: For every relational structure H the problem CSP(H) is either solvable in polynomial time or is NP-complete.

While a wide variety of methods based on model theory, database theory, graph homomorphisms, etc. have been used to attack the dichotomy conjecture, it was the discovery of the algebraic approach by Jeavons et al. [11] (see also [2]) that played the decisive role in resolving the conjecture. Using the algebraic approach the dichotomy conjecture was confirmed in a number of important cases, see, e.g., [1, 3, 4, 5, 10]. A recent collection of surveys on intermediate results on the complexity of the CSP and its variants, as well as, other applications of the algebraic approach can be found in [12]. Finally, in 2017 the Bulatov and Zhuk independently confirmed the dichotomy conjecture [6, 15] for arbitrary finite relational structures. For a less technical presentation of the first of these results see [7].

In the first half of this tutorial we survey the connections of the CSP and the dichotomy conjecture with other areas of logic and computer science, and outline the history of the problem. In the second half we explain the main ideas of the first of the solution algorithms [6, 15] for nonuniform CSPs and provide some examples.

References


What is the number of integer points in a convex polytope? This problem is of great interest in combinatorics and discrete geometry, with many important applications ranging from integer programming to statistics. From computational point of view it is hopeless in many dimensions, as the knapsack problem is a special case. Perhaps surprisingly, in bounded dimension the problem becomes tractable. How far can one go? Can one count points in projections of polytopes, finite intersections of such projections, etc?

We will survey both classical and recent results on the problem, emphasizing both algorithmic and complexity aspects. Some elegant hardness results will make an appearance in dimension as small as three. If time permits, we will discuss connections to Presburger Arithmetic and decidability problems for irrational polyhedra.

Joint work with Danny Nguyen.
This talk concerns the well-studied model of “cake-cutting” for studying questions regarding notions of fair division of resources. We focus on discrete versions rather than using infinite-precision real values, specifically, focusing on the communication complexity. Using general discrete simulations of classical infinite-precision protocols (Robertson-Webb and moving-knife), we roughly partition the various fair-allocation problems into 3 classes: “easy” (constant number of rounds of logarithmic many bits), “medium” (poly-log total communication), and “hard”.

Our main technical result concerns two of the “medium” problems (perfect allocation for 2 players and equitable allocation for any number of players) which we prove are not in the “easy” class. Our main open problem is to separate the “hard” from the “medium” classes.

Joint work with Simina Brânzei
Every hard computational problem has some hard instances, that form a “hard core” of the problem. In this talk I will discuss if the hard cores do exist, how easy is to sample hard cores, how one can measure “hardness” of the instances, etc. Our approach is based on Challenger-Solver games, introduced some time ago by Yuri Gurevich. A modification of the famous Golod-Shafarevich construction of infinite periodic groups provides first examples of natural easily samplable “really hard” cores. This leads to a discussion of computational hardness of search problems and descriptive complexity.
For the past 40 years computer scientists generally believed that NP-complete problems are intractable. In particular, Boolean satisfiability (SAT), as a paradigmatic automated-reasoning problem, has been considered to be intractable. Over the past 20 years, however, there has been a quiet, but dramatic, revolution, and very large SAT instances are now being solved routinely as part of software and hardware design. In this talk I will review this amazing development and show how automated reasoning is now an industrial reality.

I will then describe how we can leverage SAT solving to accomplish other automated-reasoning tasks. Sampling uniformly at random satisfying truth assignments of a given Boolean formula or counting the number of such assignments are both fundamental computational problems in computer science with applications in software testing, software synthesis, machine learning, personalized learning, and more. While the theory of these problems has been thoroughly investigated since the 1980s, approximation algorithms developed by theoreticians do not scale up to industrial-sized instances. Algorithms used by the industry offer better scalability, but give up certain correctness guarantees to achieve scalability. We describe a novel approach, based on universal hashing and Satisfiability Modulo Theory, that scales to formulas with hundreds of thousands of variables without giving up correctness guarantees.
TALKS
Let $\Gamma$ be a finite simple graph. The graph group $G_\Gamma$ (also known as partially commutative group) is the group such that the vertex set of $\Gamma$ serves as the generator set, and relations are as follows: two generators $x$ and $y$ commute (i.e. $[x, y] = 1$) if and only if the vertices $x$ and $y$ are connected by an edge in the graph $\Gamma$. Graph groups can be defined in various varieties of groups, for example, in the variety of nilpotent groups, solvable groups, etc. Two groups are called universally equivalent if the sets of all universal sentences which hold on these groups coincide. Five years earlier, the author together with A.A. Mishchenko proved a criterion for universal equivalence of two nilpotent graph groups. The resulting criterion was complicated and computationally hard. In the talk a new convenient approach to the description of universally equivalent nilpotent graph groups will be proposed. The new approach based on the notion of closed subset of vertices of $\Gamma$, compression of $\Gamma$ and lattice of closed subsets of $\Gamma$. 

For a given graph $F$, let $v(F)$ be the number of its vertices, $I(F)$ denote the class of all graphs containing a copy of $F$ as an induced subgraph, and $D[F]$ denote the minimum quantifier depth of a sentence in first-order logic with adjacency and equality relations that defines $I(F)$. It is obvious that $D[F] \leq v(F)$. In [1] it has been shown that $D[F] = 4$ for all graphs $F$ on 4 vertices except for paw graph $K_3 + e$ and its complement, for which $D[F] = 3$.

Our contribution is the following. For all $\ell > 4$, we give an example of graph $F$ on $\ell$ vertices such that $D[F] = \ell - 1$. Also, we prove that if all of the components of $F$ are pairwise isomorphic complete multipartite graphs, then $D[F] = \ell$. Finally, we show that, for every $F$ on 5 vertices, $D[F] \geq 4$.

References


(joint work with E.D. Kudryavtsev, A.S. Shlychkova, M.E. Zhukovskii)
We study asymptotic behaviour of the first order properties (properties expressible in first order logic) of binomial random graphs $G(n, p)$. We say that the random graph $G(n, p)$ obeys the Zero-One Law if for each first-order graph property its probability tends to 0 or tends to 1. We also say that the random graph $G(n, p)$ obeys the Zero-One $k$-Law if for each first-order graph property with quantifier depth no more than $k$ its probability tends to 0 or tends to 1.

Let $\alpha \in (0, 1)$ be a real number. It was proven by J. Spencer and S. Shelah in 1988 that the Zero-One Law holds for $G(n, n^{-\alpha})$ if and only if $\alpha$ is irrational. We say that the rational $\alpha$ is in $k$-spectrum if the random graph $G(n, n^{-\alpha})$ does not obey the Zero-One $k$-Law.

In 2012, M. Zhukovskii proved that the smallest number in $k$-spectrum is $\frac{1}{k-2}$. The full structure of $k$-spectrum remains unexplained. It is known, however (J. Spencer, 1990), that 14-spectrum is infinite. Moreover, M. Zhukovskii proved that $\frac{1}{2}$ is the limiting point of 5-spectrum while 3-spectrum is finite. Finally, it was proven by A. Matushkin and M. Zhukovskii in 2018 that there can be no limiting points in 4-spectrum but $\frac{1}{2}$ and $\frac{3}{5}$.

Our result is the following. We prove that $\frac{1}{2}$ and $\frac{3}{5}$ are not the limiting points in 4-spectrum. Thus, we conclude that the 4-spectrum is finite. Therefore, the minimal $k$ such that $k$-spectrum is infinite is 5.

References


We consider the decidability problems of the universal and existential theories of the class of all partially ordered sets in a language without constants.

A partially ordered set (poset) is an algebraic structure $\mathcal{P} = \langle P \mid L \rangle$ with domain set $P$ and language $L = (\leq (P))$, where $\leq$ - partial order predicate. On $\mathcal{P}$ three axioms hold:

1. $\forall p \in P \ p \leq p$ (reflexivity);
2. $\forall p_1, p_2 \in P \ p_1 \leq p_2 \land p_2 \leq p_1 \rightarrow p_1 = p_2$ (antisymmetry);
3. $\forall p_1, p_2, p_3 \in P \ p_1 \leq p_2 \land p_2 \leq p_3 \rightarrow p_1 \leq p_3$ (transitivity).

Atomic formula of language $L$ in variables $X$ is either equality $x_i = x_j$ or the inequality $x_i \leq x_j$, where $x_i, x_j \in X$.

Formula of language $L$ is an expression that is defined recursively as follows:

1. Atomic formula is a formula;
2. If $\varphi$ is a formula, then $\neg \varphi$ is also a formula;
3. If $\varphi, \psi$ are formulas, then $\varphi \lor \psi$ and $\varphi \land \psi$ are also formulas;
4. If $\varphi$ is a formula, then $\exists x_i \varphi$ and $\forall x_i \varphi$ are also formulas.

Sentence of language $L$ is a formula without free variables. Universal sentence is a sentence of following form $\forall x_1 \ldots \forall x_n \Psi(x_1, \ldots, x_n)$, where $\Psi$ is a formula in $n$ free variables. The existential sentence is defined similarly.

The class of posets of language $L$ is denoted by $\mathcal{P}$. Elementary theory of class $\mathcal{P}$ is set $Th(\mathcal{P})$ of all sentences of the language $L$ that is true in all posets in $\mathcal{P}$. The universal theory of the class $\mathcal{P}$ is the subset $Th_\forall(\mathcal{P}) \subseteq Th(\mathcal{P})$ of universal sentences of the elementary theory of class $\mathcal{P}$. The existential theory of the class $\mathcal{P}$ is defined similarly.

Theory $T$ of posets class $\mathcal{P}$ of language $L$ is decidable if there is an algorithm that checks whether $\varphi$ in theory $T$ or not for any sentence $\varphi$ in language $L$.

Now, we formulate the main results of the work.

**Theorem 1.** The universal theory of the class of all posets in a language $L$ without constants is decidable.

**Theorem 2.** The existential theory of the class of all posets in the language $L$ without constants is decidable.

Also, results were obtained that characterize the complexity of these problems.

**Theorem 3.** The problem of decidability of the existential theory of a class of posets is NP-hard.

**Theorem 4.** The problem of decidability of the universal theory of a class of posets is co-NP-hard.
We apply algebraic approach to the study of algorithms of hierarchical clustering. Moreover, we discuss the outlier detection by such algorithms and prove that almost all datasets contain outliers relative to the given data generation procedure. Also, we show the connection between the last result and the zero-one law in monadic second-order logic for linear orders.

Let us explain our results more carefully.

The clustering is an important problem of data analysis. Let us consider a general scheme of agglomerative clustering process. In the beginning, each element is in a cluster of its own. The clusters are then sequentially combined into larger clusters, until all elements end up being in the same cluster. At each step, the two clusters separated by the shortest distance are merged. In our research we will use the following cluster distance function

\[ d(A, B) = \min\{d(a, b) \mid a \in A, b \in B\} \] (single-linkage clustering).

The result of any agglomerative clustering can be visualized as a dendrogram, which shows the sequence of cluster fusions.

Thus, the process of agglomerative clustering starts with the distance matrix of a data. Since the singly-linkage clustering use order (not values) on matrix entries, we propose to generate \( n \times n \)-matrices as follows:

1. randomly generate the upper triangle of a matrix by elements of linearly ordered set of \( n(n - 1)/2 \) elements;
2. write zeros on the diagonal;
3. reflect the upper triangle to the lower one.

Thus, we have exactly \( (n(n - 1)/2)! \) distance \( n \times n \)-matrices.

For the given procedure of matrix generation, one can prove the following.

**Theorem 1.1.** Let \( D(M) \) be the dendrogram defined by the agglomerative clustering for a distance \( n \times n \)-matrix \( M \). Then the probability \( \Pr(D(M) \text{ has a leaf adjacent to the tree root}) \) tends to 1 with \( n \to \infty \).

*The author was supported by Russian Science Foundation (project 18-71-10028)
One can apply this theorem to the problem of outlier detection. We say that an element $x$ is an outlier if the agglomerative clustering puts $x$ into a leaf adjacent to the dendrogram root. In other words, $x$ is very far from other data elements $x_1, x_2, \ldots, x_n$, and we have two clusters $\{x_1, x_2, \ldots, x_n\}, \{x\}$ before the last cluster fusion.

Hence, the last theorem states that almost all datasets contain outliers with respect to the given procedure of data generation.

There are connections between the obtained result and the famous zero-one law in model theory. One can prove that the property "a dataset contains an outlier" may be written as a sentence of the monadic second-order logic over linear orders. Above we show this sentence is almost surely true. Probably, it is a consequence of the more general result

Conjecture. Does the zero-one law hold for linear orders in monadic second-order logic?
We study the Diophantine problem (decidability of systems of equations) in different families of solvable groups. We show that for any group $G$ in each of these families there exists a ring of algebraic integers $O$ that is interpretable in $G$ by systems of equations. This reduces the Diophantine problem of $O$—conjectured undecidable—to the same problem in $G$, and it leads us to conjecture that the Diophantine problem in $G$ is undecidable. The families where such result is obtained include all finitely generated non-virtually abelian nilpotent groups and all polycyclic groups that are not virtually metabelian. Note that the Diophantine problem of virtually abelian groups has long been known to be decidable (their first-order theory is). We also prove undecidability of the Diophantine problem in free solvable groups and in ‘most’ nilpotent groups by studying asymptotic properties of random nilpotent groups.

Joint work with Alexei Miasnikov and Denis Ovchinnikov
We consider a $\text{Cay}(\mathbb{Z}_2^n, s)$ - a Cayley graph over boolean cube group, there generating set contains all vectors with exactly $s$ non zero coordinates. Each vertex is associated with computing node in network with synchronized time, exactly one node is marked as a leader. Vertexes $b$ is called antipodal to $a$ iff each automorphism which has $a$ as fixed point also have $b$ as fixed point. The task is to find out the vertex, that is antipodal to the leader (we consider range of parameters, there each vertex has exactly one antipodal vertex). We present fast protocol that solves this problem: it stops in $O\left(\frac{n}{s}\right)$ steps, at each step at most $O(\log n)$ bits transferred through each edge.
The Quantified Constraint Satisfaction Problem $QCSP(\Gamma)$ is the problem to evaluate a sentence of the form $\forall x_1 \exists y_1 \ldots \forall x_n \exists y_n (R_1(...) \land \ldots \land R_s(...))$, where $R_1, \ldots, R_s$ are relations from the constraint language $\Gamma$. This problem is a generalization of the Constraint Satisfaction Problem, where only existential quantifiers are allowed.

We study the complexity of $QCSP(\Gamma)$ for different constraint languages on finite sets. Unlike the Constraint Satisfaction Problem, where the problem for every constraint language is either tractable, or NP-complete, $QCSP(\Gamma)$ can be PSpace-complete. It was conjectured by Hubie Chen that $QCSP(\Gamma)$ is either tractable, or NP-complete, or PSpace-complete.

We disproved this conjecture and showed that for some constraint languages $\Gamma$ the problem $QCSP(\Gamma)$ can be coNP-complete, DP-complete and so on.

Also, we characterized the complexity of the Quantified Constraint Satisfaction Problem for constraint languages on 3-element domain containing all unary singleton relations (so called idempotent case), i.e. we showed that for such languages $QCSP(\Gamma)$ is either tractable, or NP-complete, or coNP-complete, or PSpace-complete.
The Constraint Satisfaction Problem provides a common framework for expressing many common computational tasks from diverse areas of computer science. In its purest form, CSP with a fixed constraint language on a finite set, it is amenable to a computational complexity classification via polymorphisms (higher-arity symmetries of solution spaces), and abstract algebraic objects capturing their properties. An extensive research program exploring this connection to universal algebra culminated in resolving the so-called algebraic CSP dichotomy conjecture.

Recently, Brakensiek and Guruswami proposed a substantial generalization of the CSP, called Promise CSP, which is motivated by open questions about the approximability of variants of satisfiability and graph colouring. They demonstrated that the basic algebraic approach can be extended to this new setting and successfully employed to study computational complexity of Promise CSPs.

Similarly to the CSP, to fully exploit the algebraic approach we must lift it from concrete properties of polymorphisms to their abstract properties. In this talk, I will introduce the abstract polymorphism theory for Promise CSPs and demonstrate how it can be applied to improve the state-of-the-art in approximate graph coloring by showing that it is NP-hard to find a 5-coloring of a given 3-colorable graph. This is joint work with Andrei Krokhin and Jakub Oprášal from Durham University.
While, in general, the isomorphism problem for finitely presented groups is undecidable, for virtually free groups it has been shown to be decidable by Krstić. Later the complexity has been improved to primitive recursive by Sénizergues in the case that the input is either given as two context free grammars for the word problems or as finite extensions of free groups.

Here, we present an algorithm for the following problem: given a context-free grammar for the word problem of a virtually free group $G$, compute a finite graph of groups with finite vertex groups and fundamental group $G$. Our algorithm is non-deterministic and runs in doubly exponential time. Using Krstić’s algorithm for testing fundamental groups of graphs of groups for isomorphism, it follows that the isomorphism problem of context-free groups can be solved in doubly exponential space.

Moreover, if, instead of a grammar, a finite extension of a free group is given as input, the construction of the graph of groups is in NP and, consequently, the isomorphism problem in PSPACE.

While the algorithm itself is rather simple and uses only standard tools from formal language theory, its proof of correctness is more geometric and makes use of the structure tree theory introduced by Dicks and Dunwoody.

This talk is based on joint work with Géraud Sénizergues.
Universal algebraic geometry is a relatively new direction in mathematical research. The results in algebraic geometry over specific groups, rings, algebras, graphs, etc., can be called the practical part of this branch of mathematics, many articles and even books are devoted to them. The theoretical part of universal algebraic geometry is a new section of the model theory, within the framework of which the basic algebraic-geometric concepts are introduced, problems are posed and results are proved for an arbitrary algebraic structure of an arbitrary language. In the monograph by E. Yu. Daniyarova, A. G. Myasnikov, V. N. Remeslennikov “Algebraic geometry over algebraic structures”, Novosibirsk: Publ. SB RAS, 2016, 243 p., the theoretical part of universal algebraic geometry is introduced. The main ideas of this monograph will be presented on the talk. For clarity and ease of understanding, algebraic-geometric concepts and solutions of algebraic-geometric problems will be described in the projection to the category of abelian groups. In the talk,

- there will be defined equations, algebraic sets, irreducible algebraic sets, radicals, coordinate groups over abelian groups;
- it will be shown how all coordinate groups, irreducible coordinate groups, all algebraic sets, irreducible algebraic sets over abelian groups are classified;
- there will be described geometrically equivalent abelian groups and universally geometrically equivalent abelian groups;
- it will be shown that among all special algebraic-geometric classes in the category of abelian groups, only the class of equational co-domains is of interest, and also the description of abelian groups from this class will be given.
We consider some problems of aggregation of individual preferences. We show that under rather general assumptions there are only two clones of aggregation rules that allow invariant symmetric classes of preferences, each of these clones being generated by a single function.

Let $A$ be a finite set and $r$ a natural number. The symbol $[A]^r$ denotes the set of all $r$-element subsets of $A$. Individual preferences are modeled by $r$-choice functions on a set $A$, i.e., functions $c : [A]^r \to A$ satisfying $f(p) \in p$ for any $p \in [A]^r$. The set of all $r$-choice functions on a set $A$ is denoted by $C_r(A)$. A set $\mathcal{D} \subseteq C_r(A)$ is called symmetric if $c \in \mathcal{D} \Rightarrow c_\sigma \in \mathcal{D}$ for any permutation $\sigma$ of $A$ where $c_\sigma(p) = \sigma^{-1}(c(\sigma(p)))$ for any $p \in [A]^r$. A (simple local) aggregation rule is a function $f : A_{\leq r}^n \to A$ where $A_{\leq r}^n = \{ a \in A^n \mid \text{ran } a \leq r \}$, see [1] (cf [2]). For all $c_1, c_2, \ldots, c_n \in C_r(A)$ and $f : A_{\leq r}^n \to A$ the symbol $f(c_1, c_2, \ldots, c_n)$ denotes the $r$-choice function $c$ defined by $c(p) = f(c_1(p), c_2(p), \ldots, c_n(p))$ for all $p \in [A]^r$. An aggregation rule $f : A_{\leq r}^n \to A$ preserves a set $\mathcal{D} \subseteq C_r(A)$ if $f(c_1, c_2, \ldots, c_n) \in \mathcal{D}$ for all $c_1, c_2, \ldots, c_n \in \mathcal{D}$. The Galois connection generated in natural sense by the preservation relation is denoted $(\text{Inv}_r, \text{Pol}_r)$. A set $\mathcal{D} \subseteq C_r(A)$ has the Arrow property if $\text{Pol}_r(\mathcal{D})$ contains only projections (dictatorship rules). All symmetric sets without the Arrow property were classified in [3], see also [1]. In addition, it is shown [4] that if $r = 2$ (this is the most important case) and $|A| \geq 5$ then for any set $\mathcal{D} \subseteq C_r(A)$ without the Arrow property the set $\text{Pol}_r(\mathcal{D})$ consists of functions generated by the “counting-out game” function $\ell$ defined by $\ell(x, y, y) = \ell(y, x, y) = \ell(y, y, x) = \ell$. This result can be considered as a generalization of Arrow’s impossibility theorem [5]. In essence, it means that there are no acceptable aggregation rules for symmetric sets of preferences.

For positive results, we consider a more general situation. A set $\mathcal{D} \subseteq C_r(A)$ is called trivial if $\mathcal{D} = \{ c \in C_r(A) : c \upharpoonright_B = d \upharpoonright_B \}$ for some $d \in C_r(A)$ and $B \subseteq [A]^r$ (a trivial set $\mathcal{D}$ is preserved by any aggregation rule). A set $\mathcal{D} \subseteq \mathcal{P}(C_r(A))$ is called trivial if it contains only trivial sets. A set $\mathcal{D} \subseteq \mathcal{P}(C_r(A))$ is called symmetric if $\mathcal{D} \subseteq \mathcal{D} = \mathcal{D} \subseteq \mathcal{D}$ for any permutation $\sigma$ of $A$ where $\mathcal{D}_\sigma = \{ c_\sigma : c \in \mathcal{D} \}$ (for example, the class of all single-peaked domain [6] is symmetric). Let $\partial : A_{\leq 2}^3 \to A$ be a majority function. We prove the following dichotomy theorems.

**Theorem 1.** Let $|A| \geq 5$. Let $f : A_{\leq 2}^n \to A$ be a non-dictatorship aggregation rule and $\mathcal{D} \subseteq \text{Inv}_2(f)$ a non-trivial symmetric set. Then $\mathcal{D} \subseteq \text{Inv}_2(\partial)$ or $\mathcal{D} \subseteq \text{Inv}_2(\ell)$.
Let $f : A_{\leq 2}^n \rightarrow A$ and $C \subseteq C_r(A)$. A set $D$ is compatible with the pair $(f, C)$ if $D \subseteq C$ and $f(c_1, c_2, \ldots, c_n) \in C$ for all $c_1, c_2, \ldots, c_n \in D$. A set of all sets what is compatible with $(f, C)$ is denoted by $\text{Comp}(f, C)$. A function $c \in C_r(A)$ is called rational if $c(p) = \max_p p$ for some linear order $\succ$ on $A$. The set of all rational function $c \in C_r(A)$ is denoted by $\mathcal{R}_r(A)$.

Theorem 2. Let $|A| \geq 5$. Let $f : A_{\leq 2}^n \rightarrow A$ be a non-dictatorship aggregation rule and $C \subseteq \text{Comp}(f, \mathcal{R}_2(A))$ a non-trivial symmetric set. Then there is a symmetric class $D \subseteq \mathcal{P}(\mathcal{R}_2(A))$ such that

1. $D \subseteq \text{Inv}_2(\partial) \cap \text{Inv}_2(f)$ or $D \subseteq \text{Inv}_2(\partial) \cap \text{Inv}_2(f)$ and

2. For all $C \in C$ there is $D \in D$ such that $C \subseteq D$.

References


For the interpretations of linear orderings in the standard model of Presburger arithmetic \((\mathbb{N}, +)\), we show that a Hausdorff-style rank can be used as a necessary condition of multi-dimensional definability. We constructively establish a connection between the interpretability properties of an order and its Cantor condensation and, furthermore, conjecture a criterion of Presburger definability of linear orderings for all dimensions, proving it for \(m = 2\).

Classification of linear orderings according to complexity of their representation encompasses a number of interconnected problems and approaches. This question is naturally connected with the various notions of rank going back to Hausdorff’s seminal work [3]. Khoussainov, Rubin & Stephan [4] develop the theory of Kantor-Bendixson-style rank for linear orderings based on the operation of condensation, which fuses the points at a final distance between them. One could notice that all orderings that are trivialized after a, maybe transfinite, number of such iterations, are scattered (SLO). We give an overview of the results obtained (partially jointly with F. Pakhomov) in classifying the scattered linear orderings according to their definability in Presburger arithmetic, that is, by a formula with (standard) order and addition, without multiplication.

The question of interpreting Presburger arithmetic in its own finitely axiomatizable subtheories naturally leads to the behaviour of the implied interpretation of its order relation, which further generalizes into the study of all linear orderings interpretable in Presburger arithmetic in some number of dimensions. Using a modification of rank description, the following was established [7]:

**Theorem 1** (A. Zapryagaev, F. Pakhomov, 2017) All linear orderings \(m\)-dimensionally interpretable in Presburger arithmetic have a rank \(m\) or below.

Note that this gives only a necessary condition, as for the trivial argument of cardinality, there are more scattered linear orderings of any rank \(\geq 2\) than formulas in Presburger signature. However, the following holds nevertheless:

**Theorem 2** (A. Zapryagaev, 2018) Let \(\iota : L \rightarrow \mathbb{N}\) be an \(m\)-dimensional interpretation of an SLO \(L\) in the standard model of arithmetic \((\mathbb{N}, +)\). Then its condensation \(cL\) possessed an \(m - 1\)-dimensional interpretation.
Furthermore, this interpretation of $cL$ can be extracted from a particular interpretation of $L$ through a canonical construction.

In order to establish this, one requires to match to each Presburger-definable set (see also [5]) its dimension, akin to [2], which expresses the maximal number of linearly independent vectors generating some sublattice in it, based on .

Using this, the it is possible to establish:

**Theorem 3** (A. Zapryagaev, 2019) Let $ι: L → N$ be a 2-dimensional interpretation of an SLO $L$ in the standard model of arithmetic $(N,+)$. Then there is some $n ∈ [2; 4]$ that $L$ is a limiting of the lexicographical ordering of $Z^n$ onto some Presburger-definable set.

The conjecture is that it holds for each $n ≥ 1$. This construction gives a possibility to create a lambda calculus producing all definable orderings. The problem can be further generalized to automatic linear orderings which, per [1], are exactly ones interpretable in Presburger arithmetic with one additional predicate, and which still do not possess an exact explicit necessary and sufficient condition.

**References**


The time required for pathogen recognition by the immune system is critical for the organism survival. For viral diseases, recognition is performed by clones of T-lymphocytes with different types of receptors. During maturation, those T-lymphocyte clones are selected, which receptors have high dissociation constants with most of the body’s epitopes (short protein fragments). Viral proteins for which there was no such selection can be detected by the unexpectedly low dissociation constant of their epitopes with receptors [1].

Each individual receptor is contacting thousands of epitopes. The time of each contact is determined by the dissociation constant only on average. According to the developed ideas, signal transmission through the membrane occurs with as great probability, as large the clusters consisting of laterally related receptors it has [2].

It is shown that an increase in the duration of lateral bonds in the vicinity of the receptor in contact with the epitope makes possible the nucleation of macroclusters formation. Each cluster of receptors is assigned a value (playing the role of entropy) representing the degree of disorder of lateral bonds in it at a particular point in time [3]. The cluster entropy changes statistically insignificantly for clusters smaller than the critical size. The development of supercritical clusters is characterized by an increase in entropy. The probability of supercritical clusters formation is calculated as a function of the dissociation constant of pathogenic epitopes.

REFERENCES

Modelling of multiphase flow in porous media of oil and gas natural reservoirs is a very important part of modern design of development of oil and gas fields. All decisions in development are made on the basis of comprehensive multivariate set of models. It is crucial to estimate filtration (reservoir) parameters for these models - absolute permeability, capillary pressure and relative permeability. These filtration parameters can be estimated experimentally but it is expensive, difficult to recreate reservoir conditions, difficult to provide sufficient number of experiments. Pore-scale modelling of flow in porous media along with digital core analysis can solve these problems what is briefly described in the report.

Solving of the mentioned above problems depends on time spent for numerical calculations and computational recourses needed for pore-scale modelling because multivariate sets of calculations have to be done. This report briefly describes modelling of one- and two-phase flow in network models of pores and capillaries (pore network models [3]) as a part of pore-scale modelling. Pore network modelling has optimal combinations of detalization and simplicity of porous media models, calculation time and computational recourses what is important for fast multivariate calculations.

This report is intended to show used systems of difference equations for pore network modelling and their specific aspects. It shows comparisons (specific aspects of applications, calculation time, computational complexity and so on) of different methods for solving of this kind of systems of equations: the algebraic multigrid method (AMG), the biconjugate gradient stabilized method (BiCGStab), the conjugate gradient method with the symmetric successive over-relaxation (CG-SSOR), the method with using of continuous symmetry groups and others. Applications of the developed method on the basis of continuous symmetry groups [4], [5] are presented for the considered systems using obtained results of group classifications of similar to pore network models difference schemes of equation of gas flow in porous media. The results of group classification of difference schemes have been obtained using results from [1], [2].

Obtained results of comparison of numerical methods can be applied in problems of filtration parameters estimations of core samples for oil and gas natural reservoirs. It can save calculation time and computational recourses what is important in practical problems because of increasing number of used model variants and their detalization.
Acknowledgments. The reported study was funded by RFBR according to the research project 16-29-15119.

References


Russian Workshop on Complexity and Model Theory

Abstracts

June 9–11, 2019          MIPT          Moscow, Russia