NII Shonan Meeting on: Parameterized Complexity and the Understanding, Design and Analysis of Heuristics

Gregory Gutin, Kazuo Iwama, Dimitrios M. Thilikos

NII Shonan Meeting Report

No. 2013-2

Parameterized Complexity and the Understanding, Design, and Analysis of Heuristics

Gregory Z. Gutin
Kazuo Iwama
Dimitrios M. Thilikos

May 06–11, 2013

Parameterized Complexity and the Understanding, Design, and Analysis of Heuristicst

Organizers:
Gregory Z. Gutin (Royal Holloway, University of London)
Kazuo Iwama (Kyoto University)
Dimitrios M. Thilikos (University of Athens)

May 06–11, 2013

In the parameterized/multivariate framework, NP-hardness is just the beginning: a result about the null-parameterization. What follows is a rich dialogue between theory and practice, with the goals of:

- explaining the effectiveness of established heuristics and

- designing better heuristics in mathematically disciplined ways

The workshop brought together researchers from both universities and industry, who are interested in exploring how a multivariate view of complexity analysis and algorithm design can lead to industrially useful algorithms in new, mathematically systematic ways.

The vast majority of practical problems are NP-hard and thus we cannot hope to obtain algorithms that compute exact or optimal solutions efficiently, at least according to the one-dimensional (n, the input size) worst-case framework that essentially contrasts worst-case running times that are polynomial in n, versus worst-case running times that are exponential in n.

Recent discussions of the future of theoretical computer science have pointed to the importance of heuristic algorithms for NP-hard problems. In 2009, in his plenary talk at the FOCS Theory Day, Richard Karp posed the question

> *"why do heuristics work so well in practice?"*

as one of the most important challenges in the theory of computing. Applied computing communities have been aware of this issue for decades, and have lamented the relative mission-failure of theoretical computer science. Theoretical computer science, has reciprocally lamented its lack of support, both by core funding agencies, and in Computer Science curriculums. There is a central dysfunction between theory and practice to which Karp points.

In real life, typical inputs to a computational problem are the outputs of other resource-constrained processes, both natural and artificial. While datasets may be large, significant (although often multifaceted) structure is almost always present and worth taking into account in algorithm design and complexity analysis.

Parameterized complexity attends to the underlying structure of real-world input distributions, as well as structure in the computational objectives, such as size of solutions, noise in the data, goodness of the desired approximation, or combinations of these. (See [1, 2, 4].) Secondary measurements (that may form an aggregate parameter) capture such crucially relevant problem structure and open up the possibility of exploiting this structure in order to design efficient worstcase algorithms — or to identify algorithmic strategies that can usefully be incorportated into practical heuristics. The central notion of parameterized complexity is a multivariate generalization of polynomial time that is termed fixed-parameter tractability (FPT), meaning solvability in time $f(k) + n^c$, where f is some function of the parameter k, n is the input size, and c is a constant — in other words, the input size contributes polynomially to the complexity, and any exponential costs are confined to some function of the parameter k.

Parameterized algorithms, both directly and indirectly, have made major contributions to practical computing in such areas as Computational Biology and (increasingly) Artificial Intelligence. The historical record is muddy. Since computing began, applications-determined implementors have of course attended to relevant secondary measurements that allowed their implementations to succeed on real-world data-sets (but without any attention to theoretical frameworks). Meanwhile, onedimensionally framed and trained theorists have crept towards a multivariate outlook, under the rhetoric of exact algorithmics.

Although the historical record is muddied, it is essentially the case that the modern computational revolutions in Biology and Medicine would not have happened without parameterized/multivariate algorithmics. Explicit mathematical formulation of the issues adds power to the algorithm engineering opportunities, not only in these application domains, but wherever NP-hardness of central computational tasks has been identified, and heuristics are currently the main game.

Real-life input distributions have significant structure, but what is it, and how should it be explicitly declared in realistic multivariate complexity bookkeeping? The multivariate approach to complexity analysis and algorithm design often demands significant dialog between theory results, and practical computing experience, including the adaptation of FPT algorithms and mathematically-centered algorithmic strategies into practical-computing-targeted heuristics:

- It is well-known that the input of every FPT parameterized problem can be pre-processed, in polynomial time, such that the reduced instance (losing no information) has size bounded (only) by a function of the parameter (for many NP-hard problems, a polynomial function). Such kernelization subroutines can be highly useful in practical computing, regardless of how the reduced instances are solved. For the first time: a useful mathematical theory of pre-processing.

- Parameterization also offers, for the first time, a useful theory of the complexity of local search subroutines.

- FPT subroutines are also useful in the "genetic recombination" phase of hybrid (e.g., memetic) meta-heuristics.

- Algorithmic strategies for FPT algorithms, such as iterative compression and greedy localization, can be systematically adapted into theoretically

organized classes of heuristics (see [Karp11]) that are competitive or superior to previous efforts, or in some cases, exactly the same: the successful heuristic was simply an unrecognized FPT algorithm, according to an unrecognized but relevant parameter (see examples in [DF99]).

- The mathematical determination that a parameterized problem is FPT often involves estimates concerning mathematical relationships that are far from the plausible but unproven truth of the matter, or worst-case assumptions about unparameterized aspects of the input that are far from typical. Systematic algorithm engineering compromises about such issues can lead to novel and effective heuristics derived from FPT algorithms (see [TM09] for a pioneering example).

- The central innovation of the parameterized approach to algorithms and complexity is, of course, simply the explicitly multivariate view of complexity analysis and algorithm design. Beyond that, at the core of the multivariate algorithmics project, there is a richer vision of the workflow between theory and practice. In the classical framework of P versus NP, a problem is NP-hard, and that is almost the end of the story. As Karp wrote in [Karp11]:

    "Except in rare cases, the problems are NP-hard, and the performance guarantees provided by polynomial-time approximation algorithms are far too pessimistic to be useful."

In the multivariate workflow there are two complementary creative principles:

Deconstruction of hardness results. Most hardness proofs in both classical and parameterized complexity are unreasonable. Read the proofs and ask why the constructions are unreasonable for real-world inputs. This often points to realistic FPT parameterizations.

Model enrichment. When FPT results are obtained, it is often productive to enrich the definition of the problem, in order to obtain better traction in real-world applications. Parameters can be complex aggregates that wrap onto problem realism.

Because a single classical problem can be parameterized in an unlimited number of ways, this rich mathematical dialog exposes many opportunities for fresh approaches in the design of practical algorithms and heuristics. The goal of the Workshop was to explore, systematize and programmatically articulate these opportunities.

References

[1] Downey, R.G. and Fellows, M.R.: Parameterized complexity, Springer, 1999.

[2] Flum, J. and Grohe, M.: Parameterized Complexity Theory, Springer, 2006.

[3] Karp, R.: "Heuristic algorithms in computational molecular biology," J. Computer and System Sciences 77 (2011), 122-128.

[4] Niedermeier, R.: An Invitation to Parameterized Algorithms,

[5] Tazari, S. andM¨uller-Hannemann, M.: "Dealing with Large Hidden Constants: Engineering a Planar Steiner Tree PTAS", proceedings of ALENEX 2009, pp. 120–131.X 2009, pp. 120–131.

Overview of Talks

Practical Aspects of Fixed-Parameter Algorithms: An Implementations' Viewpoint

Faisal Abu Khzam, Lebanese American University

The talk will cover some implementation techniques and a discussion of design alternatives, including a comparison between best-worst-case algorithms and algorithms based on simple heuristic priorities. Some parallel implementation strategies will be discussed as well.

The Matrix Representation with Flipping problem and the FlipCut heuristic

Sebastian Boecker, University of Jena, Germany

In computational phylogenetics, supertree methods assemble phylogenetic trees with non-identical but overlapping taxon sets, into a larger supertree: These supertrees contains all taxa of all input trees and describes the evolutionary relationship of these taxa. This problem can be formalized in different ways, to cope with contradictory information in the input. I will concentrate on the Matrix Representation with Flipping (MRF) framework: The input trees are encoded in a 0/1/?-matrix, and we search for a minimum set of 0/1-flips such that the resulting matrix admits a directed perfect phylogeny. This is an NP-hard problem, but if all input trees share the same set of taxa, the problem is fixed-parameter tractable. Otherwise, the problem is W[2]-hard and cannot be approximated within a constant factor. I will present a data reduction as well as an ILP formulation for the problem which, unfortunately, both do not work well in practice. Finally, I will present a novel heuristic for the problem, named FlipCut supertrees. Evaluations of the heuristic are very promising.

Cluster Editing in Practice

Sebastian Boecker, University of Jena, Germany

Cluster Editing ask to transform an undirected graph into a disjoint union of cliques using a minimum number k of edge modifications. I will talk about practical issues that concern the implementation of algorithms for Cluster Editing, including the growth of the parameter k, running times in practice, best-performing reduction rules, and others.

Combinatorial optimization using diversities

David Bryant, University of Otago

A *diversity* is a generalisation of metric spaces to beyond pairwise comparisons. Formally, it is a pair (X, δ) where X is a set and δ is defined on finite subsets of X satisfying the two axioms

1. $\delta(A) = 0 \Leftrightarrow |A| \leq 1$;

2. $\delta(A \cup C) \leq \delta(A \cup B) + \delta(B \cup C)$ whenever $B \neq \emptyset$.

Note that δ restricted to pairs gives a metric. Examples of diversities include the diameter, the length of the minimum tour, and the length of an optimal Steiner tree. Diversities were introduced in a recent paper of Bryant and Tupper, and generate a rich theory which we are only just beginning to explore. We survey some of the main results and then speculate wildly on their use in applied combinatorial optimization and data analysis. In particular we discuss their role in hypergraph algorithms and show how L_1 embedding strategies for graphs and metrics generalize to analogous results for hypergraphs and diversities (at least that is the plan).

Exploring parameterised complexity in computational topology

Benjamin Burton, The University of Queensland

Topological decision problems in three dimensions are notoriously difficult: the mere existence of an algorithm is often a major result, and many important algorithms have yet to be studied from a complexity viewpoint. Even "simple" problems, such as unknot recognition (testing whether a loop of string is knotted), or 3-sphere recognition (testing whether a triangulated 3-manifold is topologically trivial), have best-known algorithms that are worst-case exponential time. In practice, however, some of these algorithms run surprisingly well in experimental settings, and we discuss some reasons why parameterised complexity now looks to be the "right" tool to explain this behaviour. We also present some initial forays into the parameterised complexity of topological problems, including both fixed-parameter-tractability and W[P]-hardness results.

What have I been doing recently in parameterized complexity

Rod Downey, Victoria University of Wellington

I will mention some of the work in this area I have been looking at. This includes the new book with Mike, plus lots of new projects with Judith Egan, Mike, Fran and Peter Shaw. These concern permissive search, incremental computation, hueristics, and parameterizing by the number of numbers. I will give a brief overview of these.

Inferring Haplotypes from Genotypes on a Pedigree

Patricia Evans, University of New Brunswick

In genetics, it is considerably easier to determine genotype information (what values an individual has at each of a number of marker sites) than it is to determine haplotype information (how those values are split between the chromosomes). Since haplotype information can be very useful in pedigree analysis, we would like to be able to infer haplotypes from genotypes for a set of individuals with known family relationships in a pedigree. This work examines how haplotypes including recombination can be inferred as an application of Edge Bipartitization. This is joint work with Peter Doan.

Parameterized k-opt

Fedor V. Fomin, Univ. of Bergen

We discuss results on the parameterized complexity about searching in the k-exchange neighborhoods for the following optimization problems
   - Problems on planar graphs
   - Feedback Arc Set in Tournaments (FAST)
   - Chordal completion (fill-in)

Domination Analysis of Heuristics

Gregory Gutin, Royal Holloway, University of London

Many heuristics do not provide approximation guarantee in the worst case and so we cannot compare worst case performance of heuristics for the same problem using approximation. Domination analysis is an alternative way of comparing heuristics based on the concept of domination number proposed in 1996 by Glover and Punnen. (Interestingly, some results in the area were obtained already in the 1970s.) Consider the Asymmetric Travelling Salesman problem (defined on weighted complete directed graph). The domination number of an ATSP heuristic H donated $\mathsf{dom}(H, n)$ is the maximum integer d such that for every instance of ATSP on n vertices H finds a solution not worse than d solutions. For example, for every $n > 1$ there are instances on n vertices for which the nearest neighbour heuristic (NN) will find the unique worst possible solution. Thus, the domination number of NN is 1. There are ATSP heuristics with domination number at least $(n - 2)!$ and even higher. We'll concentrate mainly on TSP heuristics, but heuristics for other problems will also be mentioned.

Exponential Neighborhoods

Gregory Gutin, Royal Holloway University of London

Exponential neighbourhoods contain exponential number of solutions. However, there are some such neighbourhoods for which the best solution can be found in polynomial time. We'll consider such neighbourhoods for TSP and other problems. We should be able to get something much bigger if we replace

'polynomial time' with 'FPT time', but for which problems we can do it and what is the parameter?

Implementing and testing fixed-parameter algorithms

Falk Hüfner, Technische Universität Berlin

The talk is about various aspects of implementing and testing fixed-parameter algorithms, such as ensuring correctness, debugging, speedups, and performance evaluation and visualization.

Parameterized Testability

Kazuo Iwama, School of Informatics, Kyoto University

In graph property testing, given a graph represented by an oracle, we want to test whether the graph satisfies some predetermined property P or ϵ-far from satisfying the property. Roughly speaking, a graph is called ϵ-far from a property if we need to modify an ϵ-fraction of the graph to make it satisfy the property. A most important focus in this topic has been on general characterizations for (constant-time) testable properties, which have obtained much progress in the last decade: For the dense graph model, Alon et al. finally obtained a purely combinatorial necessary and sufficient condition for P to be testable that is closely related to the Szemerédi Regularity Lemma. For the bounded-degree graph model, due to Benjamini et al., any property is testable if it is minor-closed. In this talk, we are still interested in a general framework of testable properties, but our approach is a bit different from the qualitative one as above. Rather, we are interested in "parameterized properties" that are quite common in NP optimization problems. A bit curiously, this direction has not been so popular in the context of property testing. The obvious reason is that the problem becomes less interesting in both the dense graph model and the bounded degree model. For instance, consider the problem k-VERTEX COVER for a constant k. In the dense graph model, we want to decide whether a graph has a vertex cover of size at most k or we need to remove at least ϵn^2 edges to have the property. It turns out that this property is easily testable just by accepting only graphs with at most $\epsilon \times n^2$ edges. To avoid this triviality, we consider the general graph model with an augmentation of random edge sampling capability. It is shown that a variety of such problems, including k-VERTEX COVER, k-FEEDBACK VERTEX SET, k-MULTICUT, k-PATH FREE and k-DOMINATING SET, are constant-time testable if k is constant. It should be noted that the first four problems are fixed parameter tractable (FPT) and it turns out that algorithmic techniques for their FPT algorithms (bounded-branching tree search, color coding, etc.) are also useful for our testers. k-DOMINATING SET is W[2]-hard, but we can still test the property in constant time since the definition of ϵ-farness makes the problem trivial for non- sparse graphs that are the source of hardness for the original optimization problem. We also consider k-ODD CYCLE TRANSVERSAL, which is another well-known FPT problem, but we only give a sublinear-time tester when k is a constant. This is a joint work with Yuichi Yoshida.

LP-Based Parameterized Algorithms for Separation Problems

Daniel Lokshtanov, Univ. Bergen

We show that the most basic branching algorithm for Vertex Cover turbo-boosted by a classical linear-programming based reduction rule runs in time $4^{k-OPT(LP)}\mathsf{poly}(n)$, where $OPT(LP)$ is the value of the optimum solution to the LP relaxation of the problem. It turns out that this algorithm can be used to solve Odd Cycle Transversal and Almost 2-Sat, as well as a number of other problems, in time $4^k\mathsf{poly}(n)$. We then give an improved algorithm for Vertex Cover with running time $2.32^{k-OPT(LP)}\mathsf{poly}(n)$, this improves considerably over the currently fastest parameterized algorithms for Odd Cycle Transversal and Almost 2-SAT.

Linear kernel via conflict packing – application to FAST and dense RTI problems

Christophe Paul, CNRS-LIRMM

We introduce a new technique to design kernelization algorithms for parameterized problems, namely Conflict Packing. We illustrate this technique on two well-studied problems: FAST and RTI. For the former, one is given a tournament $T = (V, A)$ and seeks a set of at most k arcs whose reversal in T leads to an acyclic tournament. While a linear vertex-kernel is already known for this problem, it requires a constant-factor approximation algorithm as a subroutine. Using the Conflict Packing, we show how to avoid this subroutine to find a so-called safe partition. In the RTI problem, one is given a set of vertices V and a dense collection R of rooted binary trees over three vertices of V and seeks a rooted tree over V containing all but at most k triplets from R. Using again the Conflict Packing, we prove that the dense RTI problem admits a linear vertex-kernel. This result improves the on the quadratic known kernel. joint work with A. Perez and S. Thomassé

Implementing Courcelle's Theorem

Peter Rossmanith

I will make a demonstration of our MSO solver and explain how it works. The solver is based on game trees.

Tractable Parameterizations for the Minimum Linear Arrangement Problem

Hadas Shachnai, Technion, Israel

The Minimum Linear Arrangement (MLA) problem asks to embed a given graph on the integer line so that the sum of the edge lengths of the embedded graph is minimized. Most layout problems are either intractable, or not known

to be tractable, parameterized by the treewidth of the input graphs. We investigate MLA with respect to three parameters that provide more structure than treewidth. In particular, we give a factor $1 + \epsilon$ -approximation algorithm for MLA parameterized by (ϵ, k), where k is the vertex cover number of the input graph. By a similar approach, we describe two FPT algorithms that exactly solve MLA parameterized by, respectively, the max-leaf and clique-cover numbers of the input graph. Joint work with Michael Fellows, Danny Hermelin and Frances Rosamond.

Solving Hard Problems Incrementaly

Peter Shaw, Charles Darwin University

By utilizing an incremental (or solution repair) formulation, some hard problems, when formulated with appropriate parameters, become FPT. This talk describes how algorithmic solutions of such problems can be applied as effective FPT subroutines to improve many local search problems. To illustrate this we consider the Dominating Set problem. We show that the incremental Dominating set problem is FPT. A number of formulations of this problem were considered, but most of these resulted in hard problems. We show a FPT algorithm for one particular formulation and further prove that this problem has no polynomial kernel. This is a joint work with Rod Downey, Judith Egan, Michael Fellows, and Frances Rosamond.

Some Algorithmic Challenges in Arc Routing

Manuel Sorge, Technische Universität Berlin

Arc Routing problems occur naturally in many logistic tasks like delivering mail, garbage disposal, street sweeping and snow plowing. Such a task is formulated, for example, in the Chinese Postman problem where, given an edge-weighted graph, one searches for a minimum-weight tour that visits all edges in the graph. In this simple form, Chinese Postman is solvable in polynomial time, but many variants that arise in applications yield NP-hard problems. Yet, despite the practical relevance, the literature on parameterized complexity of Arc Routing problems is rather sparse. We give a brief survey of the known (parameterized) complexity results and identify some interesting open questions. In particular, we briefly report on our theoretical and empirical studies related to the Rural Postman problem. This is joint work with René van Bevern, Rolf Niedermeier, and Mathias Weller.

Deconstructing Algorithms: From Theory to Engineering

Siamak Tazari, Google

We present the generic idea of deconstructing a sophisticated theoretical algorithm, and engineering its parts to obtain a useful, practical "heuristic" via a case study on Steiner tree in planar graphs. Afterwards, we will briefly talk about the disconnect between theoretical computer science and the state of the industry today and propose possible directions on bringing them closer together.

From parameterized complexity to heuristics

Dimitrios M. Thilikos, CNRS-LIRMM & University of Athens

In this talk we describe some of the the existing links netween parameterized complexity and heuristics. More particular we present how the notion of kernelization can serve as a formal mathematical framework for studying preprocessing and we explain why the multivariate computational complexity viewpoint is necessary for this approach. We also give several research directions for the study of various heuristic approaches using tools and techniqeus from parameterized algorithm design.

FPT for Scalable Mining Algorithms

Takeaki Uno, National Institute of Informatics

Data mining algorithms often enumerates patterns/structures included in the given data, for exhaust search of candidates. They usually deal with data that are noisy but have some property (sparse, locally dense, having hubs, etc.) The time complexity of the algorithm are usually exponential in the input size, but practically they are efficient. One of the reason is that, the number of solutions is not exponential, relatively small, and the algorithms are output sensitive (# solutions (resp., # edges) is a kind of "fixed parameter"). We would like to introduce fixed parameters to data mining/enumeration algorithms so that we can solve difficult problems, and can understand the mechanism of the practical efficiency.

Efficient Enumeration of the Directed Binary Perfect Phylogenies from Incomplete Data

Yoshio Okamoto, Japan Advanced Institute of Science And Technology

We study a character-based phylogeny reconstruction problem when an incomplete set of data is given. More specifically, we consider the situation under the directed perfect phylogeny assumption with binary characters in which for some species the states of some characters are missing. Our main object is to give an efficient algorithm to enumerate (or list) all perfect phylogenies that can be obtained when the missing entries are completed. While a simple branch-and-bound algorithm (B&B) shows a theoretically good performance, we propose another approach based on a zero-suppressed binary decision diagram (ZDD). Experimental results on randomly generated data exhibit that the ZDD approach outperforms B&B. We also prove that counting the number of phylogenetic trees consistent with a given data is #P-complete, thus providing an evidence that an efficient random sampling seems hard. Joint work with Masashi Kiyomi and Toshiki Saitoh.

Graph Isomorphism Problem Parameterized by Width Parameters

Yota Otachi, Japan Advanced Institute of Science And Technology

We study the parameterized complexity of the graph isomorphism problem

when parameterized by width parameters related to strong tree decompositions. We first present FPT algorithms for some width parameters, and then provide a unified explanation for various isomorphism results concerned with parameters related to tree decompositions. As a first step towards intractability results for parameterized graph isomorphism we develop an FPT Turing-reduction from strong tree width to the a priori unrelated parameter maximum degree. This is joint work with Pascal Schweitzer.

List of participants

- Dr. Faisal Abu Khzam, Lebanese American University
- Dr. Sebastian Boecker, Friedrich-Schiller University Jena
- Dr. David Bryant, University of Otago
- Professor Benjamin Burton, The University of Queensland
- Professor Rod Downey, Victoria University of Wellington
- Dr. Patricia Evans, University of New Brunswick
- Professor Rudolf Fleischer, German Univ Technology of Oman
- Professor Fedor V. Fomin, Univ. of Bergen
- Prof. Tobias Friedrich, University of Jena
- Professor Gregory Z. Gutin, Royal Holloway, University of London
- Dr. Falk Hüffner, Technische Universität Berlin
- Professor Kazuo Iwama, Department of Communication and Computer Engineering, Kyoto University
- Professor Ken-ichi Kawarabayashi, National Institute of Informatics
- Dr. Daniel Lokshtanov, Univ. of Bergen
- Professor/Dr. Matthias Müller-Hannemann, Martin-Luther-University Halle-Wittenberg, Halle, Germany
- Dr. Frank Neumann, The University of Adelaide
- Professor Christophe Paul CNRS – LIRMM
- Professor Peter Rossmanith, RWTH Aachen University
- Professor Saket Saurabh, The Institute of Mathematical Sciences
- Professor Hadas Shachnai, Technion, Haifa
- Dr. Peter Shaw, Charles Darwin Univ
- PhD student Manuel Sorge, TU Berlin
- Dr. Siamak Tazari, Google
- Professor, Dimitrios M. Thilikos, LIRMM Montpellier & University of Athens,
- Dr. Takeaki Uno, National Institute of Informatics (NII)
- Professor Vladimir Estivill-Castro, Griffith University
- Dr. Yoshio Okamoto, University of Electro-Communications
- Assistant Professor Yota Otachi, Japan Advanced Institute of Science and Technology.

Photo

Foto of participants of Shonan Seminar 018, Wednesday, May 8, 2013