

Multivariate Investigation of NP-Hard Problems: Boundaries Between Parameterized Tractability and Intractability

Uéverton dos Santos Souza¹, Fábio Protti¹, Maise Dantas da Silva¹, Dieter Rautenbach²

¹Programa de Pós-Graduação em Computação
Universidade Federal Fluminense, Niterói, Brasil

²Institut für Optimierung und Operations Research
Universität Ulm, Ulm, Germany

{usouza, fabio}@ic.uff.br, maisedantas@id.uff.br,
dieter.rautenbach@uni-ulm.de

Abstract. *In this thesis we present a multivariate investigation of the complexity of some NP-hard problems, i.e., we first develop a systematic complexity analysis of these problems, defining its subproblems and mapping which one belongs to each side of an “imaginary boundary” between polynomial time solvability and intractability. After that, we analyze which sets of aspects of these problems are sources of their intractability, that is, subsets of aspects for which there exists an algorithm to solve the associated problem, whose non-polynomial time complexity is purely a function of those sets. Thus, we use classical and parameterized complexity in an alternate and complementary approach, to show which subproblems of the given problems are NP-hard and latter to diagnose for which sets of parameters the problems are fixed-parameter tractable, or in FPT. This thesis exhibits a classical and parameterized complexity analysis of different groups of NP-hard problems. The addressed problems are divided into four groups of distinct nature, in the context of data structures, combinatorial games, and graph theory: (I) and/or graph solution and its variants; (II) flooding-filling games; (III) problems on P_3 -convexity; (IV) problems on induced matchings.*

1. Introduction

The question “ $P = NP?$ ” is the most important open question in computer science, and the theory of NP-completeness was developed to show which problems probably do not have polynomial-time algorithms. Though it is nice to know that some problems do not have polynomial time algorithms unless $P = NP$, many NP-hard and NP-complete problems must still be solved in practice (especially those with real-world applications); therefore it is natural to ask if each of these problems admits an algorithm whose non-polynomial time complexity is purely a function of some subset of its aspects (in practice many aspects of the problem often has bounded size or value). Questions about the existence of such algorithms are addressed within the theory of parameterized computational complexity developed by Downey and Fellows [1, 3, 4].

In the first part of the thesis, we present a detailed review of the theory of parameterized complexity, where the concepts and techniques involving fixed-parameter tractability and intractability are discussed.

The parameterized complexity theory became very popular in recent years and has become an important research topic at many universities around the world. In this

sense, the Brazilian community can be somewhat outdated, with few Brazilian researchers working on this area. A major contribution of this thesis is to collaborate in popularizing the parameterized complexity nationwide. In this regard, as a consequence of our work, we can mention a narrowing of the relationship between researchers and the professors Frances Rosamond and Michael R. Fellows, one of the authors of the parameterized complexity. This relationship resulted in some studies developed with their co-authorship, and in particular in their visit to Brazil in November 2014 and their subsequent participation as invited speakers in the 6th Latin American Workshop on Cliques in Graphs.

In addition, one of the goals of this thesis it is to make an analysis on the sources of polynomial time intractability of some interesting problems. The problems studied are of distinct nature, and the concepts discussed have applications in many areas such as software engineering, distributed systems, artificial intelligence, bioinformatics, operational research, social networks, automation, game theory, among others. Moreover, the proofs presented are of several types such as NP-hardness proofs, polynomial algorithms, structural characterizations, W[1]-hardness and W[2]-hardness proofs, FPT algorithms, polynomial kernel results, and infeasibility of polynomial kernels.

Below we present the list of papers developed and published throughout my PhD, related to this thesis.

1. Tractability and hardness of flood-filling games on trees.
Theoretical Computer Science, v. 576, p. 102-116, 2015.
2. Maximum Induced Matchings close to Maximum Matchings.
Theoretical Computer Science, 2015. (to appear)
3. Complexity Properties of Complementary Prism.
Journal of Combinatorial Optimization, 2015. (to appear)
4. An algorithmic analysis of Flood-it and Free-Flood-it on graph powers.
Discrete Mathematics and Theoretical Computer Science, v. 16, p. 279-290, 2014.
5. Revisiting the complexity of and/or graph solution.
Journal of Computer and System Sciences, v. 79, p. 1156-1163, 2013.
6. Complexidade Parametrizada para Problemas em Grafos E/OU.
Pesquisa Operacional para o Desenvolvimento, v. 4, p. 160-174, 2012.
7. The Flood-It game parameterized by the vertex cover number.
Electronic Notes in Discrete Mathematics, LAGOS 2015.
8. On Graphs with Induced Matching Number Almost Equal to Matching Number.
Electronic Notes in Discrete Mathematics, LAGOS 2015.
9. On P3-convexity of Graphs with Bounded Degree.
Lecture Notes in Computer Science, v. 8546, p. 263-274, 2014.
10th Conference on Algorithmic Aspects of Information and Management (AAIM).
10. Parameterized Complexity of Flood-Filling Games on Trees.
Lecture Notes in Computer Science, v. 7936, p. 531-542, 2013.
19th International Computing and Combinatorics Conference (COCOON).
11. Complexity of Geodetic Number Problem in Graphs with Maximum Degree 4.
13th Cologne-Twente Workshop on Graphs and Comb. Optimization (CTW), 2015.
12. The P3-Convexity in the Complementary Prism of a Graph
13th Cologne-Twente Workshop on Graphs and Comb. Optimization (CTW), 2015.
13. Parameterized And/Or Graph Solution.
12th Cologne-Twente Workshop on Graphs and Comb. Optimization (CTW), 2013.

14. Optimal Variability Selection in Product Line Engineering.¹
24th Conference on Software Eng. and Knowledge Engineering (SEKE), 2012.
15. Inundação em Grafos.
16th Congreso Latino Iberoamericano de Investigación Operativa (CLAIO), 2012.

Other projects were developed and submitted along my PhD, such as:

16. P_3 -Convexity Problems on Bounded-Degree and Planar Graphs.
Theoretical Computer Science.
17. Tractability and Kernelization Lower Bound for And/Or Graph Solution.
Discrete Applied Mathematics.
18. Complexity of Geodetic Number Problem in Graphs with Bounded Degree.
Journal of Computer and System Sciences.
19. A Multivariate Investigation of Flood-It Game. *Discrete Applied Mathematics*.
20. Parameterized Problems on Complementary Prisms. *Discrete Mathematics*.

The main results of this work are briefly presented as follows.

2. Complexity of And/Or Graph Solution

The first group of studied problems involve two important data structures used for modeling many real-world applications, *and/or graphs* and *x-y graphs*. An and/or graph is an acyclic, edge-weighted directed graph containing a single source vertex, such that every vertex $v \in V(G)$ has a label $f(v) \in \{\text{and}, \text{or}\}$. X-y graphs are a generalization of and/or graphs: every vertex v_i of an x-y graph has a label x_i-y_i meaning that v_i depends on x_i of its y_i out-neighbors. We investigate the complexity of finding a solution subgraph H of such digraphs, which must contain the source and obey the following rule: if a vertex is included in H then x_i of its out-edges must also be included in H , where an and-vertex has $x_i = y_i$, and an or-vertex has $x_i = 1$. The optimization problems associated with such graphs are formally defined as MIN-AND/OR and MIN-X-Y, respectively. We denote by MIN-AND/OR(k) and MIN-X-Y(k) its classical parameterized versions.

Theorem 1 MIN-AND/OR remains NP-hard even for a very restricted family of and/or graphs where edges have weight one and or-vertices have out-degree at most two.

Theorem 2

(a) The parameterized problem MIN-AND/OR⁰(k), whose domain includes and/or graphs allowing zero-weight edges, is W[2]-hard.

(b) MIN-X-Y(k) is W[1]-hard.

(c) MIN-AND/OR(k) is fixed-parameter tractable.

Using a framework developed by Bodlaender et al. (2009) and Fortnow and Santhanam (2011), based upon the notion of compositionality, we show that the above parameterized problem does not admit a polynomial kernel unless $NP \subseteq coNP/poly$.

Theorem 3 MIN-AND/OR(k) has no polynomial kernel unless $NP \subseteq coNP/poly$ and consequently $PH \subseteq \Sigma_3^p$.

3. Flooding-filling games

The second group of studied problems consists of variants of a one-player combinatorial game known as the Flood-Filling Game, which is played on a colored board and whose objective is to make the board monochromatic (“flood the board”) with the minimum

¹In this paper the authors apply and/or graphs in software engineering problems.

number of flood moves. A flood move consists of assigning a new color c_i to the a pivot tile p and also to all the tiles connected to p by a monochromatic path immediately before the move. The flood-filling game where all moves use the same pivot p is denoted by Flood-It. When the player can freely choose which tile will be the pivot of each move the game is denoted by Free-Flood-It.

A complete mapping of the complexity of flood-filling games on trees is made, charting the consequences of single and aggregate parameterizations. Furthermore, we show that Flood-It on trees and Restricted Shortest Common Supersequence (RSCS) are analogous problems, which proves some FPT and W[1]-hard of cases of Flood-it. In addition, we prove that Flood-It remains NP-hard when played on 3-colored trees, which closes an open question. We also present a general framework for reducibility from Flood-It to Free-Flood-It. Analyzing the behavior of these games when played on other classes of boards, we describe polynomial time algorithms, and some NP-hard cases. Finally, we show that Flood-it is fixed-parameter tractable when parameterized by the vertex cover number, and it has a polynomial kernel if the number of colors is a second parameter.

Let Π be a flood-filling game and let $S = \{s_1, \dots, s_n\}$ be a subset of the aspects of Π . $[S_1]\text{-}\Pi(S_2)$ is the family of parameterized problems where the aspects in $S_1 \subseteq S$ are fixed constants and the aspects in $S_2 \subseteq S$ are aggregate parameters.

We consider the following aspects of the problem: c - number of colors; λ - number of moves; d - maximum distance of the pivot; o - maximum orbit; k - number of leaves; r - number of bad moves, $r = (\lambda - c)$.

Theorem 4

- (a) $[d]\text{-Flood-It}$ on trees remains NP-hard when $d = 2$.
- (b) $[d]\text{-Flood-It}(c)$ is in FPT and admits a polynomial kernelization.

Theorem 5

- (a) Flood-It on trees and RSCS are analogous problems.
- (b) Flood-It(c, k, λ) on trees is p -analogous to RSCS($|\Sigma|, \ell, \Lambda$).

We remark that the book [2] (see Appendix) reports that the SCS problem is solvable in polynomial time when each string has size two. This assertion is false, as we can see by Theorems 4 and 5.

Corollary 1 *Flood-It(k, c) on trees is W[1]-hard.*

A colored rooted tree is a pc-tree (phylogenetic colored tree) if no color occurs more than once in any path from the root to a leaf. A pc-tree T is a cpc-tree if each color occurs exactly once in any path from the root to a leaf. Restricting attention to phylogenetic colored trees, we have significant effects on problem complexity.

Theorem 6

- (a) Flood-It(k) on pc-trees with pivot root is W[1]-hard.
- (b) Flood-It on trees remains NP-hard even when restricted to cpc-trees.
- (c) Flood-It(r) on cpc-trees with pivot root is in FPT.
- (d) Flood-It(k, r) on trees is in FPT.

Let G be a graph, $v \in V(G)$, and ℓ a positive integer. The graph $\psi(G, v, \ell)$ is constructed as follows: (i) create ℓ disjoint copies G_1, \dots, G_ℓ of G ; (ii) contract the copies v_1, v_2, \dots, v_ℓ of v into a single vertex v^* . Let \mathcal{F} be a class of graphs. Then:

$$\psi(\mathcal{F}) = \{G \mid G = \psi(G', v, \ell) \text{ for some triple } (G' \in \mathcal{F}, v \in V(G'), \ell > 0) \}.$$

Theorem 7 *Flood-It on \mathcal{F} is reducible in polynomial time to Free-Flood-It on $\psi(\mathcal{F})$.*

Theorem 8

- (a) *Flood-It on trees remains NP-hard even restricted to 3-colored trees.*
- (b) *Free-Flood-It on trees remains NP-hard even restricted to 3-colored trees..*

Theorem 9 *In Free-Flood-It on pc-trees, there always exists an optimal free-flooding which is a flooding with pivot root.*

Theorem 10

- (a) *Flood-it is solvable in polynomial time on P_n^2 , C_n^2 , and $2 \times n$ circular grids.*
- (b) *Free-Flood-it remains NP-hard on P_n^2 , C_n^2 , and $2 \times n$ circular grids.*

Theorem 11 *Flood-it on graphs is fixed-parameter tractable, FPT, when parameterized by the size of the minimum vertex cover (k).*

Theorem 12 *Flood-it on graphs admits a polynomial kernelization when parameterized by the size of the minimum vertex cover (k) and the number of colors (c).*

4. P_3 -convexity

The third group comprises some problems on P_3 -convexity. More specifically we are interested in identifying either the minimum P_3 -geodetic set or the minimum P_3 -hull set S of a graph, from which the whole vertex set of G is obtained either after one or eventual iterations, respectively. Each iteration adds to a set S of vertices all the vertices of $V(G) \setminus S$ with two neighbors in S .

One of the most elementary models of the spreading of a property within a network – like sharing an idea or disseminating a virus – one can consider a graph G , a set S of vertices of G that initially possesses the property, and an iterative process whereby new vertices u are added to S whenever sufficiently many (usually two) neighbors of u are already in S . Similar models were studied in various contexts, such as statistical physics, social networks, marketing, and distributed computing.

Theorem 13 *Let c be a positive integer. If G is a graph with $\delta(G) \geq \frac{|V(G)|}{c}$, then*

$$h_{P_3}(G) \leq 2 \left\lceil \frac{\log(2c)}{\log\left(\frac{2c^2}{2c^2-1}\right)} \right\rceil + 2c^3.$$

Corollary 2 *A minimum P_3 -hull set of a graph G with $\delta(G) \geq \frac{|V(G)|}{c}$ (for some constant c) can be found in polynomial time.*

Theorem 14

- (a) *P_3 -HULL NUMBER remains NP-complete on planar graphs G with $\Delta(G) = 3$.*
- (b) *A minimum P_3 -hull set of a cubic graph can be found in polynomial time.*

Theorem 15

- (a) *P_3 -GEODETIC NUMBER remains NP-complete on planar graphs G with $\Delta(G) = 3$.*
- (b) *P_3 -GEODETIC SET(k) is W[2]-hard.*
- (c) *P_3 -GEODETIC SET(k, Δ) is fixed-parameter tractable.*

The complementary prism $G\bar{G}$ of G arises from the disjoint union of the graph G and its complement \bar{G} by adding the edges of a perfect matching joining pairs of corresponding vertices of G and \bar{G} .

Theorem 16

- (a) *To decide whether a compl. prism $G\bar{G}$ has a P_3 -geodetic set of size k is NP-complete.*
- (b) *A minimum P_3 -hull set of a compl. prism $G\bar{G}$ can be found in polynomial time.*

5. Problems on induced matchings

The last group of problems studied in this thesis focus on a classical topic in graph theory. These problems are related to maximum matchings, maximum induced matchings, and the distance between them in a graph. The matching number $\nu(G)$ of G is the maximum cardinality of a matching in G , and a matching with $\nu(G)$ edges is a maximum matching of G . An induced matching is a set M' of edges of G at pairwise distance at least 2. The induced matching number $\nu_2(G)$ of G is the maximum cardinality of an induced matching in G , and an induced matching with $\nu_2(G)$ edges is a maximum induced matching. The distance between a maximum matching of a graph G and its maximum induced matching is the difference between the cardinality of these sets ($\nu(G) - \nu_2(G)$). We study graphs G with $\nu(G) - \nu_2(G) \leq k$, for some non-negative integer k .

Let G be a graph with $\nu(G) - \nu_2(G) \leq k$ for some non-negative integer k .

Let M_1 and M_2 be a maximum matching and a maximum induced matching of G , respectively, and let $H = (V(G), M_1 \Delta M_2)$.

If some component of H is a path with 2 edges, say $e_1 \in M_1$ and $e_2 \in M_2$, then $(M_1 \setminus \{e_1\}) \cup \{e_2\}$ is a maximum matching of G having more edges in common with M_2 than M_1 . Iteratively applying this exchange operation to M_1 , we may assume that no component of H is a path with 2 edges.

Lemma 1 *The components of H are*

- *isolated vertices,*
- *paths of length 1 whose edge belongs to $M_1 \setminus M_2$, and*
- *paths of length 3 whose two leaf edges belong to $M_1 \setminus M_2$ and whose middle edges belong to $M_2 \setminus M_1$.*

Furthermore, H has exactly $\nu(G) - \nu_2(G)$ non-trivial components.

Theorem 17 *For a fixed non-negative integer k , the graphs G with $\nu(G) - \nu_2(G) \leq k$ can be recognized in polynomial time.*

Our next result states that the recognition of those graphs G with $\nu(G) - \nu_2(G) \leq k$ that are of bounded maximum degree is fixed parameter tractable when parameterized by k .

Theorem 18 *Let Δ and k be positive integers. The graphs G with $\nu(G) - \nu_2(G) \leq k$ of maximum degree at most Δ can be recognized in $f(k, \Delta)n^c$ time where the constant c does not depend on Δ or k .*

References

- [1] R. G. Downey, M. Fellows. *Parameterized complexity*, Springer, 1999.
- [2] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*. Freeman, 1979.
- [3] J. Flum, M. Grohe. *Parameterized complexity theory*, Springer, 2006.
- [4] R. Niedermeier. *Invitation to fixed-parameter algorithms*, Oxford University Press, 2006.