Review of Doctoral Theses by Dušan Knop

Dr. Andreas Emil Feldmann

1 Summary of obtained results

The thesis titled "Structural Properties of Graphs and Efficient Algorithms: Problems Between Parameters" studies several NP-hard decision problems on graphs from the point-of-view of parameterized complexity. The general theme is to study parameters that characterize sparse or dense graph classes. As parameters for sparse graphs, the classical *tree-width* is compared with the newer *tree-depth* for the Minimum Length Bounded Cut problem, which is a generalization of Minimum Cut: for a designated vertex pair s, t, the aim is to find the minimum number of edges, such that removing them increases the distance between s and t to a given length L. The main result here is an FPT algorithm for the combined parameter L and the tree-width. As a corollary, the problem is also FPT for the parameter tree-depth, and an XP algorithm exists for the parameter tree-width (without L). It is also shown that the problem is W[1]-hard for parameter *path-width*. In particular, this shows a dichotomy between the parameters tree-width and tree-depth, which has not been observed for many problems yet. Additionally, it is also shown that no polynomial sized kernel exists for the combined parameter tree-width and L, and the main FPT algorithm is extendable to a generalization called Minimum Length Bounded MultiCut, for which more than two terminals need to be separated, although here the number of terminals is also a parameter.

The remainder of the thesis mostly considers parameters for dense graphs, most notably the *neighbourhood diversity*, which was introduced by Michael Lampis and generalizes the well-known *vertex cover number*. As a first problem, Target Set Selection is considered, which takes its inspiration from social networks (or maybe Conway's Game of Life): the aim is to select a subset S of vertices of size at most b, such that initially "activating" S, after a finite number of steps all vertices are activated. A vertex is activated if sufficiently many neighbours are active, according to a given threshold function. The first result is that the problem is FPT for the neighbourhood diversity parameter, if the threshold function is uniform, but W[1]-hard for the same parameter for general threshold functions. A different generalization of the vertex cover number is the *twin cover*, which is incomparable to neighbourhood diversity. For this parameter it is also shown that the problem is FPT if the threshold function is uniform. However for the *modular width* parameter, which is a generalization of both the twin cover and neighbourhood diversity, it is shown that the problem is W[1]-hard, even for the majority threshold function, which is a special case of uniform functions.

The next studied problem is Distance Constrained Labelling, which is a generalization of Graph Colouring: given an integer λ and k integers $p_1, \ldots, p_k$, the aim is to assign integer labels of value at most λ to the vertices, such that the labels differ by at least p_i if their distance is $i \leq k$. An even more general problem is Channel Assignment, in which edges have integer weights and the aim is to assign integer labels of value at most λ to the vertices, such that the labels of two adjacent vertices differ by at least the weight of their common incident edge. The main result is that Channel Assignment is FPT when combining the parameters neighbourhood diversity and the maximum edge weight, if the edge weights fulfil a certain uniformity condition with respect to the neighbourhood diversity. As a consequence, Distance Constrained Labelling is FPT when combining the neighbourhood diversity parameter with k and the maximum integer p_i. Moreover, for the stronger parameter of vertex cover number combined with the maximum edge weight, Channel Assignment is FPT, even without the uniformity condition. Finally, it is also shown that for the much more general *clique-width* parameter, Channel Assignment is paraNP-hard.

The last considered problem is Partition into H, which generalizes the Perfect Matching problem: for a fixed graph H the aim is to partition the vertices of the input graph G, such that each part induces a graph isomorphic to H. The first result for this problem is that it is expressible as an MSO_2 formula, and thus by Courcelle's Theorem is FPT for the parameter tree-width. For the modular width parameter, the problem is shown to be FPT, if the fixed graph H is a so-called prime graph. However, it is also shown that no polynomial sized kernel exists for this parameter, even if H is any graph with at least three vertices. Without the restriction of H to prime graphs, it is shown that using the stronger parameter neighbourhood diversity, Partition into H is FPT.

In light of the obtained results as summarized above, the thesis demonstrates the author's ability for creative scientific work.

2 Relevance and critique

Roughly speaking, the results of the thesis fall in the category of multivariate algorithmics, which can be described as an area of parameterized algorithms in which parameters are combined and compared, in order to obtain a more fine-grained view of the complexity landscape of the considered problems. Among others, this area has been promoted by established researchers such as for instance Mike Fellows and Rolf Niedermeier, and thus certainly has its significance in the broader field of parameterized complexity and algorithmics.

There are two main applications of parameterized algorithms, towards which the present thesis contributes to different degrees. The first more theoretical goal is to have a more general theory than the classical P vs NP point-of-view. The idea is that if a problem is in FPT for a certain parameter, then the problem is tractable (similar to the class P), while it is intractable if it is W[1]-hard (similar to NP-hardness). With this goal in mind, a natural question is what the most general set of parameters is that make a problem tractable in this sense. This question is clearly being addressed in the thesis for several problems, for instance by obtaining FPT algorithms for the neighbourhood diversity parameter if algorithms for the stronger parameter of vertex cover number are known.

The other more practical goal of parameterized complexity is to obtain efficient algorithms that can be used in practice. Here the focus is more on meaningful parameters, which tend to be small in practical applications. Even though the thesis mentions some practical applications of the used parameters (most notably for the Partition into H problem), this broader aspect of parameterized algorithms is not well represented in the obtained results.

The thesis can thus be mainly attributed to the theoretical aspects of the research area. In light of this one may wonder however, what motivated the particular choice of parameters and problems. The thesis sets out with the observation that parameters for sparse graphs, such as the tree-width, have been widely studied. The focus therefore shifts to dense graph classes and parameters for these. However the selection seems rather slim. In multivariate algorithmics, typically a whole plethora of parameters is considered in order to draw a complexity map of a specific problem. This map usually is a much larger version of the Hasse diagram found in Figure 1.1 of the thesis, and is marked with the complexity frontier at which a problem switches from being FPT to being W[1]-hard. Even for dense graphs, plenty more parameters than those considered in the thesis are studied in the literature (e.g. *clique cover, independent set number, dominating set number, Dilworth number, maximum induced matching, diameter, NLC-width, rank-width, boolean-width*, etc).

3 Further research questions

In light of the above critique, one pressing question obviously is the status of the complexity of the considered problems for all the other parameters for dense graphs. In particular, do any of the techniques used in the thesis to obtain FPT algorithms or hardness results carry over to other parameters as well, or do entirely new techniques have to be developed?

Apart from the motivation for the choice of parameters, what I missed most from the obtained results was a more modern view of parameterized complexity, in which the fine-grained complexity is also considered. It is often the case that a reduction not only shows that a problem is hard, but also that under the Exponential Time Hypothesis (ETH) a runtime lower bound can be obtained. For example the MULTICOLOURED k-CLIQUE problem is well-known to be solvable in time $n^{O(k)}$, but under ETH no $n^{o(k)}$ time algorithm exists, giving tight runtime bounds. Most of the reductions in the thesis are from exactly this problem, so what types of runtime lower bounds do they imply? Additionally, it is known that reductions from MULTICOLOURED k-CLIQUE can often easily be converted to reductions from the SUBGRAPH ISOMORPHISM problem, which can also give stronger runtime lower bounds. Can this technique be applied in the presented reductions as well? Furthermore, it is possible to obtain FPT runtime lower bounds, such as for k-VERTEX COVER, which is well-known to be solvable in $2^k \cdot n^{O(1)}$ time, while no $2^{o(k)} \cdot n^{O(1)}$ time algorithm exists under ETH. Can such lower bounds also be found for the presented FPT algorithms?

It is noteworthy that all considered problems of the thesis, with the exception of TARGET SET SELECTION, are generalizations of well-known graph problems. One question arising from this observation is what consequences the obtained results have for these more special MINIMUM CUT, GRAPH COLOURING and PERFECT MATCHING problems. The first and last of these problems are polynomially solvable. However a recent trend in parameterized complexity has been to apply techniques from this area to problems in P. Most prominently this has been done in order to prove polynomial lower bounds on runtimes under ETH. But also algorithms with improved polynomial runtimes have been shown to exist for fixed parameters (so-called *fully polynomial-time parameterized* algorithms). What are the consequences of the obtained results of the thesis for MINIMUM CUT, GRAPH COLOURING and PERFECT MATCHING? Do any of the used techniques give insights into these problems in terms of polynomial-time lower bounds or improved polynomial-time algorithms?

Finally, even though the thesis rather concerns theoretical than practical questions, it would still be interesting to know what the potential practical applications of the obtained results are. It is often said that data is sparse in practice. In that case, what are the applications for dense data sets? What real-world graphs have small neighbourhood diversity for instance?

4 Notable inaccuracies

An old saying goes that you cannot judge a book by its cover, and this is certainly also true for this thesis. Nevertheless, as fallible human beings we tend to make this mistake. For academic publications it is said that for every ten or more people reading the abstract, only one will read the introduction, while of every ten or more readers of the introduction, only one will go on to read the remainder. I think a case can be made that also for every ten that read the title, only one will read the abstract. The implication is that the importance of the title, abstract, introduction, and remainder decreases exponentially in this order. When I first glanced at the thesis the very first thing that hit my eye was a quite prominent spelling mistake in the title on the cover of the thesis. Looking at the abstract, I immediately found three further typos. Spelling mistakes in a thesis on a mathematical subject such as this one may seem like trivialities. Being a fallible human being myself however, the first thought I had was that this thesis will not be fun to read, since I extrapolated that it will most likely be full of spelling mistakes, which would make for a very bumpy read. To my relief this was not the case though. It seems that the author simply forgot to check the title and abstract for mistakes. What I fear though is that many future readers will make the same misjudgement of the thesis by glancing at the cover, thinking that this is not a serious piece of work based on their fist impression, and ignoring it for further consideration. I strongly suggest to correct the typos found in the title and abstract (I also found one in the acknowledgements, and a few more in the remainder of the thesis).

Given the importance of the introduction in any academic publication, I would like to give some further feedback on it (I here consider Chapter 1 as part of the introduction, since it introduces the studied parameters):

- Scientists have developed many more than just two tools to deal with algorithmic hardness, e.g. heuristics, average time complexity (randomization), pseudo-polynomial and mildly exponential time complexity, input restriction, or smoothed analysis, to name a few.
- An approximation scheme is not defined by a runtime that is polynomial in the input size and the approximation guarantee. This would be a *fully polynomial time* approximation scheme (FPTAS), and is a lot more restrictive than a polynomial time approximation scheme (PTAS). For the latter the runtime can be exponential in the approximation factor. In fact, when seen as a parameter, a PTAS is an XP-algorithm in this factor.
- The first description of an FPT algorithm in the third paragraph of the introduction is ambiguous. First off, how can a runtime be polynomial if it may depend arbitrarily on the parameter? For this you would need to assume that the parameter is constant. In this case however, an XP-algorithm also runs in polynomial time when allowing any computable time in the parameter. An FPT algorithm captures something more subtle than this, which is formally defined after the description. But in this sense the description does not "lead to" the formal definition.
- The first sentence of "Our Contributions" claims that the tree-width is widely known among theorists and practitioners. Out of my own experience, I would however doubt that it is well-known among practitioners.
- In Section 1.1 the phrase "unless something bad happens" is very informal, and in fact sounds quite funny to my ears. Right after this: I believe the class of problems tractable in exponential time is called EXP and not E.
- It is not explained in Chapter 1 why the presented parameters define sparse or dense graph classes. As this is a central theme of the thesis, it merits at least a citation to some explanation.
- Figure 1.1 seems to be missing the tree-depth parameter, which is quite central to the thesis and should therefore be included. The figure also has a parameter labelled "sd", which however is not defined.

Dr. Andreas Emil Feldmann,
28th July 2017