Review of the thesis: “Equivalences between Some Problems on Strings, Trees and Graphs” by Bartlomiej Dudek

Reviewer: Amir Abboud (Associate Professor; Weizmann Institute of Science)

General assessment: This thesis makes important contributions to the field of fine-grained complexity (FGC) by proving a collection of strong results regarding the computational complexity of some fundamental problems. The results were published in the top venues in the field and have received considerable attention from the community. Towards proving the results, the thesis introduces and employs multiple novel technical ideas that advance the tools that are at our disposal. Moreover, the thesis is well-written and well-organized.

Putting the results in context: Fine-grained complexity is a subfield of theoretical computer science at the intersection of computational complexity and algorithm design that aims to achieve a more refined understanding of the time complexity of the most basic and fundamental computational problems. A primary focus of this field is concerning the exact polynomial in the time complexity of problems in the class P. While there has always been an interest in designing better and better algorithms that reduce the exponent in the running time, in recent years there has been much efforts in finding matching lower bounds. Unfortunately, techniques for proving unconditional lower bounds do not exist, and so the community has settled for conditional lower bounds that are based on conjectures about the hardness of some core problems.

Most of the results of this thesis are typical FGC results in that they achieve an almost complete understanding of the complexity of important problems under assumptions about the hardness of more basic problems. For example, in the first set of results discussed below, they show that a large class of seemingly-complicated problems are equivalent in time complexity to the very basic 3SUM problem - a core problem in the field that has received extensive scrutiny. In a similar way, the second set of results achieves optimal results for complex-looking problems under an assumption about one of the most simple graph problems, namely of counting 4-cycles. Such results are extremely valuable since they redirect the attention of scientists from the complex-looking problems to the simpler problems.

3SUM and LDTs: One of the core problems in FGC is 3SUM in which we are given n numbers and are asked if there are three that sum to zero, i.e. whether there are x,y,z in the set such that x+y+z=0. This problem can be solved in quadratic time and one of the most important and productive conjectures in the field states that there is no subquadratic algorithm. Being such an important problem within the field, 3SUM has been extensively studied. Nonetheless, one of the most natural questions about it has remained open until the work in this thesis.

Specifically, the question is whether the complexity changes if we change the expression we are asking about from x+y+z=0 to something else such as x+2*y+z=10 or any a*x+b*y+c*z=d. If one thinks about this for a bit, it is natural to become convinced that the problem should be just as hard for any a,b,c,d unless they are degenerate in some way, e.g. a=0. However, proving this has been elusive for more than a decade leading the community to suspect that perhaps this is not true or perhaps it is true but not provable without introducing other assumptions.

Happily, Dudek et al. give a beautiful proof of this natural conjecture giving a complete answer to the question above. Their work was published at STOC, the most prestigious venue for results in FGC and in theoretical computer science in general.

Their proof is based on a combination of two things: (1) a thorough investigation using known tools such as color coding that distills the difficulty of the question to its essence, and (2) a brilliant solution for the most challenging case that is based on Behrend’s construction of sum-free sets which is a powerfully surprising statement in additive combinatorics that was used in various breakthroughs in computer science.

This beautiful paper has quickly become well-known in the community and multiple groups have tried to follow-up on it. One of the intriguing questions became whether the same classification can be established for 4SUM and its variants. Perhaps surprisingly, this question gave rise to additional challenges that could not be met for several years until works by Jin-Xu and Abboud-Bringmann-Fischer which were published in FOCS and invited to the special issue for best papers. This gives testimony to the interest in these questions.

4-Cycles, patterns in permutations, and quartet distance: Detecting and counting 4-cycles in graphs is one of the most basic computational questions that is closely related to the questions of detecting and counting triangles in graphs. The latter is arguably the most fundamental question in FGC and lies at the core of almost all other conjectures. While the 4-cycle appears to be different in complexity from triangle, and in some sense, it is much less understood, it is still considered a core problem that has been extensively studied by the community. In surprising and fascinating works by Dudek and Gawruchowski, the authors connect 4-cycle to two seemingly unrelated problems, one in permutations and one in trees. This significantly improves our understanding of the latter problems, and at the same time, makes the 4-cycle even more interesting as a subject of study in FGC. This result was also published in STOC.

The thesis asks a very interesting question about the complexity of counting small patterns in a given permutation. Specifically, they focus on patterns of length 4 and ask whether all such patterns can be counted in linear time. Before their work, this was conceivable, however, the community had been stuck at super-linear running time for a long time. The authors resolve this question by proving an equivalence with 4-cycle counting in graphs. An equivalence means that there are two way reductions between these problems, and in fact, both of these reductions turn out to be very interesting. The reduction from 4-cycle to the counting of a certain pattern of size 4 means that there exists such patterns that we do not expect to be able to count in linear time; this is based either on a conjecture about the complexity of 4-cycle or even on a more popular conjecture about the complexity of triangles. This is a great result. And in the other direction, the reduction to counting 4-cycles allows the authors to use a non-trivial algorithm for counting 4-cycles that uses fast matrix multiplication in order to improve the state of the art for counting patterns in permutations. This is an exciting example of how FGC can lead to better algorithms.

The third problem concerns the quartet distance between two trees, which is also about counting the number of certain small patterns in a given tree. Again, the authors resolve a longstanding question about the complexity of this important problem by connecting it to 4-cycles, giving both a new algorithm and a conditional lower bound.

Compression with top-trees Another result in the thesis concerns the efficiency of computing compression based on top-trees. Information-theoretically, one can compress trees with top-trees down to $O(n/\log n)$ but not better. However, the best compression one can get with an existing efficient algorithm produces compressions that are larger by a $\log\log n$ factor. It has been open whether the existing recent algorithm can be analyzed in a better way to show that the $\log\log n$ factor is not needed and that in fact it achieves the information-theoretically optimal bound. A result in this thesis shows a counter-example: a family of trees that cannot be compressed optimally with the existing algorithm. This is a very interesting result. It does not fall squarely within FGC since one is only showing a lower bound for a specific algorithm, leaving the possibility that a different efficient algorithm would get the optimal compression; however, it could be considered as a first step towards such as result.

Online Context-Free Grammar Recognition Context-free grammars (CFGs) are ubiquitous in computer science and the research into the computational questions involving them has a long history. Perhaps the most basic question is whether a given string can be derived from a given CFG. Research towards finding such algorithms has been unceasing since the 60's. More recently, FGC have given matching conditional lower bounds for the complexity of this question.

In this thesis, an interesting *online* version is asked: suppose the string arrives one character at a time, how fast can we solve the problem then?

In the offline-setting the complexity is n^{ω} with fast matrix multiplication, which is strongly subcubic. In the online setting, the only improvement on the simple n^3 in the literature was by logarithmic factors. In this thesis, the authors establish a reduction to the online version of matrix multiplication. For the latter problem, a more substantial speedup exist, shaving a $2^{\sqrt{\log n}}$ factor, and the authors are able to port this improvement into the online CFG recognition problem as well. This is a very nice result.