Review of the PhD thesis of Bartosz Rybicki

by Karen Aardal*

1 Introduction

The thesis of B. Rybicki contains several approximation algorithms for various facility location problems. These problems are among the core problems of combinatorial optimization. They are interesting from the theoretical as well as the practical point of view since they form mathematical structures that occur in many other problems.

Basically all facility location problems belong to the class of *NP-hard* optimization problems. If we accept the general belief that such problems cannot be solved in polynomial time, a very relevant question to ask, is whether or not we can develop an algorithm with polynomial running time that is guaranteed to give us a good quality solution. Such algorithms are referred to as *approximation algorithms*.

One of the early results in the area of approximation algorithms was precisely for a facility location problems: G. Cornuéjols, M.L. Fisher, G.L. Nemhauser (1977), Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. Management Science 23(8), 789–810. In this celebrated paper, the authors prove that the greedy algorithm is a $(1 - 1/e)$-approximation algorithm. Notice that the version Cornuéjols et al. were studying is the one where we want to *maximize* the profit of connecting clients to facilities minus the cost of opening facilities. The underlying mathematical "explanation" for the result was the fact that this version of the problem can be viewed as maximizing a submodular set function. The property of submodularity does not carry over to the more natural minimization version of the problem. This version is the one that has attracted most attention in the literature, and it remained an open problem for 20 years to obtain a constant factor approximation algorithm for this version. A constant factor algorithm cannot be obtained if we do not put restrictions on the assignment costs, but under the assumption that they are metric, Shmoys et al. (reference [25] in Rybicky's thesis) gave the first constant factor algorithm for the classical uncapacitated facility location problem. Since then, research on approximation algorithms for facility location problems literally exploded. Many of the world's best algorithms researchers have focused on this class of problems and this also means that the competition to derive new results is stiff. One of the most recognized group working

*Delft Institute of Applied Mathematics, Delft Institute of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands and Centrum Wiskunde en Informatica, Amsterdam, The Netherlands, k.i.aardal@tudelft.nl.

on this topic is the one around Jarek Byrka. It is with this as a background I judge the thesis. I think the results presented are impressive.

2 Chapters 1 and 2

These chapters are introductory to the main results presented. I think the level is appropriate and the background information sufficient in order to understand the remaining chapter. Chapter 1 is nicely written with a good overview of known results and the contributions of the thesis.

A minor remark is that I personally prefer to have all references ordered alphabetically, but the order followed here might be according to a template used for theses? If one wants too look whether or not a certain reference is cited in the thesis, then one has to look through the whole list every time.

3 Chapter 3

In this chapter an approximation algorithm for the so-called knapsack median problem is given. The knapsack median is a generalization of the classical K-median problem; instead of the cardinality constraint

$$\sum_{i \in F} y_i \leq K \,,$$

a knapsack constraint

$$\sum_{i \in F} w_i y_i \leq B \,,$$

is added. If $w_i = w$ for all $i \in F$, then the standard cardinality constraint is obtained.

For this problem Rybicki (based on paper [14]) derived an approximation algorithm with a performance guarantee that is a huge improvement of the previously best known algorithm; the algorithm by Rybicki at al. has a guarantee of 17.46 compared to the previously best known guarantee of 32. In addition, a so-called *bi-factor* algorithm is given. The "performance guarantee" then consist of two parts (a, b), where a indicates a maximum violation of feasibility, and b a guarantee on the objective value. The algorithm derived by the authors is a $(1 + \epsilon,\ 3.05)$-algorithm indicating that the knapsack budget constraint might be violated by a factor ϵ at most, and that the objective value in the worst case is 3.05 times the optimal value.

The analysis of the 17.46-algorithm is intricate. It combines and refines techniques based mainly on results by Kumar and by Swamy, but a lot of new results and insights have been provided. I cannot help wondering whether results similar to the one derived by An et al. [51] could at all be used here.

4 Chapters 4 and 5

Here, the capacitated K-median problem is considered. This is a very intractable problem, and so far no constant factor approximation algorithm is known. The only algorithms that are known with constant performance guarantee on the objective value accept bounded violation of some constraints.

These two chapters are quite technical, but the problem is notoriously difficult. I will be keen to discuss with the author if he sees any potential breakthrough possible in deriving a constant factor approximation algorithm in the "standard" sense, i.e., without any violation of constraints.

5 Chapter 6

This chapter treats one of my own favourite problems; the k-level uncapacitated facility location problem. First a refined and improved version of the algorithm of Rybicki's MSc-thesis is presented, then an extension to the version of the problem with penalties is derived. The important insight in this work is that the optimal integer solution has the structure of a forest, whereas the optimal LP-solution does not have this structure. The way the formulation is adapted so that the LP-relaxation optimum also has a forest structure is very clever. The analysis is elegant and the improvement over known results is substantial.

6 Concluding remarks

I think the thesis is of a very high level. As I pointed out in the beginning, the competition to derive new results is very harsh. Rybicki and co-authors (in the papers that are used as a basis for the thesis) have derived many breakthroughs on a large variety of facility location problems. Even if all problems are facility location problems, they are mathematically very different. For instance, the uncapacitated version is basically completely "governed" by the objective function, whereas median and capacitated versions are governed by the input to the constraints as well. The multi-level problems again have very different issues. To be able to get the deep feeling for such a broad class of problems in the short period of a PhD is commendable. I have been in the jury of around 30 PhD theses, and this thesis is clearly in the top 5-10 %.