Theory of Computation Laboratory 2
School of Computer and Communication Sciences

Lausanne, September 1, 2019

Report on the PhD thesis by Krzysztof Sornat

Krzysztof Sornats thesis "Approximation Algorithms for Multiwinner Elections and Clustering Problems" is a very impressive work with very strong results in approximation algorithms and he also presents very interesting connections between clustering problems and multiwinner elections. In what follows, I describe my opinions of the thesis with respect to results, quality, and presentation. I then give a short summary motivating my recommendation: My recommendation is that the thesis is passed with distinction. In other words, it is an outstanding thesis. My own expertise is in theoretical aspects of algorithms and, more specifically, in the use of convex relaxations for discrete optimization problems which is closely related to the results of the thesis.

Results

Krzysztof's thesis is on the well established field of approximation algorithms. Approximation algorithms give a rigorous mathematical way of designing and comparing the quality of efficient algorithms for NP-hard optimization problems. These are problems that are believed to have no efficient *exact* algorithms. The central question of the field is to understand within what factor an efficient (polynomial time) algorithm can approximate a considered optimization problem. The thesis focus on clustering problems which are central problems in combinatorial optimization. It establishes new impressive results and it makes interesting connections to multiwinner elections. I now give a brief explanation of the main results of the thesis.

Chapter 2: Ordered k-Median

The ordered k-median problem is a generalization of several classical clustering problems, such as k-center and k-median. The definition is as follows. We are given a set of facilities F, a set of clients C, a sorted weight function w, a metric distance function c between all the facilities and clients, and an integer k. The goal is to open at most k facilities such that the total "ordered" connection cost is minimized. For a fixed subset $S \subseteq F$, the connection cost of a client j in C is its distance to the closest facility in S. The ordered connection cost is then $\sum d_j w_j$, where d_j is the cost of j'th client with highest connection cost. The problems k-center and k-median is then obtained by setting $w = (1, 0, \ldots, 0)$ and $w = (1, 1, \ldots, 1)$, respectively. Both those special cases have constant-factor approximation algorithms and the main open question was if the ordered k-median problem also admits such a constant-factor approximation guarantee. Krzysztof answers this to the affirmative in his thesis. The result is obtained by building on previous techniques in a very clever way. In particular, he uses a rounding approach developed for k-median by Charikar & Li. However, to apply it to the ordered k-median problem is not easy. First, new insights have to be made and a "reduced" cost function is considered. This reduced cost function is obtained by "guessing" which for technical reasons turns out to be essential since the standard linear program was known to have an unbounded integrality gap. This is a very impressive result. To be honest, we were

School of Computer and Communication Sciences

Prof. Ola Svensson
EPFL I&C IIF THL2
Station 14
CH–1015 Lausanne
Suisse

+41 21 693 2641
+41 79 962 1165
ola.svensson@epfl.ch
http://theory.epfl.ch/osven

also working on this problem (unsuccessfully) and where very excited when it was announced. It was then subsequently presented in our reading group. The paper that is the basis of this chapter appeared in STOC, 2018 which together with FOCS is the premier venue of theoretical computer science.

Chapter 3: Harmonic k-Median and OWA k-Median

In this chapter a version of the k-median problem is studied where the distance function c is not assumed to satisfy the triangle inequality. This turns out to be important in multiwinner election settings but presents a big challenge for algorithm design: prior to this work all constant-factor approximation algorithms heavily relied on c being a metric.

The harmonic k-median problem is defined as follows. After selecting a set $S \subseteq F$ of k facilities, the cost of a client j is

$$1 \cdot c_1(S,j) + \frac{1}{2}c_2(S,j) + \frac{1}{3}c_3(S,j) + \ldots + \frac{1}{k}c_k(S,j),$$

where $c_\ell(S,j)$ denotes the distance from j to the ℓ:th closest facility in S. So in harmonic k-median, the cost of a client j depends on all opened facilities and not only the closest one as in standard k-median. The chapter presents a 2.36-approximation algorithm for this problem, which is especially impressive since c is not assumed to be a metric. The approach is a randomized rounding approach called dependent randomized rounding that was developed by Srinivisan and has been a very successful tool in clustering. However, the analysis is very original (since c is not a metric). Here, Krzysztof uses several strong negative correlation properties of the rounding to show the constant-factor approximation guarantee. All in all it is an impressive proof with many careful calculations. This chapter is based on a paper that was published in ICALP, 2018 which is the premier European conference in theoretical computer science.

Chapter 4: Minimax Approval Voting

In the minimax approval voting problem we are given a multiset $S = \{s_1, \ldots, s_n\}$ of binary strings of length m, and an integer k. The goal is to find a string $s \in \{0,1\}^m$ so as to minimize the maximum Hamming distance between s and the strings in S. The chapter gives the first polynomial time approximation scheme (PTAS) for this problem. A PTAS is an arbitrarily good approximation algorithm where we can get arbitrarily close to the optimum solution value if we are willing to spend more time. This result is less close to my own expertise. The techniques are to "compress" the multiset S into a small stable set X that still captures the important information of S. The problem can then be solved by roughly brute-force techniques. Apart from the first PTAS, Krzysztof also gives fixed parameter tractable algorithm where the parameter is the value of the optimum solution. This is useful as, if the optimum solution has value $> \log n$, then the problem can be solved by simple randomized rounding and Chernoff bounds. Indeed, Krzysztof uses this insight to give a faster PTAS. These results were published in the conferences WINE 2014, AAAI 2017, and in the journal AAAI (all excellent venues).

Quality and Presentation

The thesis is of outstanding quality and it is a pleasure to read. The results and the techniques of the thesis are great and very interesting. I am in particular impressed by the results of Chapters 2 and 3 as they are very close to my own research interests. I consider these as major contributions. It should also be said that the publication record of the papers that the thesis is based on is very impressive.

The writing of the thesis is also very good. It is great that Krzysztof gives a high level

School of Computer and Communication Sciences

Prof. Ola Svensson
EPFL I&C IIF THL2
Station 14
CH–1015 Lausanne
Suisse

+41 21 693 2641
+41 79 962 1165
ola.svensson@epfl.ch
http://theory.epfl.ch/osven

Theory of Computation Laboratory 2
School of Computer and Communication Sciences

overview of his results in Chapter 1 before going to the details. I could only find very few typos.

Summary and recommendation

Krzysztof's thesis is an outstanding contribution in algorithm design with a focus on approximation algorithms for clustering problems. I recommend that it is *accepted with distinction.* As I said, I deem it outstanding and I would consider it for awards if it was a PhD thesis of EPFL.

Sincerely yours,

Ola Svensson

School of Computer and Communication Sciences

Prof. Ola Svensson
EPFL I&C IIF THL2
Station 14
CH–1015 Lausanne
Suisse

+41 21 693 2641
+41 79 962 1165
ola.svensson@epfl.ch
http://theory.epfl.ch/osven