Shanghai University of Finance and Economics,
100 Wudong Street, Yangpu, Shanghai, China

March 25, 2022

To Whom It May Concern,

It is my great pleasure to review Jan Marcinkowski's PhD thesis. The dissertation is well-written, and the results are solid. In my opinion, the thesis is clearly above the bar of PhD evaluation and is ready to be submitted with only small changes. Please see the attached for the detailed report.

Should you have any questions regarding the review, please do not hesitate to contact me either through my work email at `<bundit@sufe.edu.cn>` or personal email at `<Lbundit@gmail.com>`.

Sincerely Yours,

Bundit Laekhanukit
Associate Professor, Shanghai University of Finance and Economics

Review of Jan Marcinkowski's PhD thesis

Summary

This thesis studies the approximability of three combinatorial optimization problems:

1. Minimum Maximal Matching (MMM): Given a (bipartite) graph, the goal is to find a "maximal" matching with minimum-cardinality (or minimum-weight).
2. Proportional Approval Voting System: Given a set of candidates chosen by voters, the goal is to (approximately) determine the set of candidates who win the election.
3. Capacitated k-Median (CkM): Given a set of points in a metric space, the goal is to find k points as centers (or facilities) so that the sum of the distances from each other point (clients) to its closest center is minimum.

The dissertation is well-written, and the results are solid. In my opinion, the thesis is clearly above the bar of PhD evaluation and is ready to be submitted with only small changes.

Minimum Maximal Matching

This problem is a variant of the classical minimum-weight "maximum" matching, where the goal is to find a "maximal" matching rather than the one with maximum cardinality. However, when the objective function changes from maximum to maximal, the known Hungarian method for the minimum-weight matching problem is no longer applicable. The minimum maximal matching problem is, indeed, known to be NP-hard.

A well-known greedy algorithm for the maximum weight matching problem yields a factor 2-approximation and gives the same factor for the minimum maximal matching problem. However, prior to the result presented in this thesis, it was not known whether a better than factor-two approximation algorithm existed for the problem.

The author of this thesis and his colleagues show a lower bound result that matches the known upper bound, thus closing the problem (albeit under the Unique Game Conjecture). The hardness result can be extended to bipartite graphs but with a weaker threshold of $4/3 - \epsilon$. The tight hardness result can be obtained under stronger complexity assumptions – the strong unique game conjecture and the small set expansion conjecture. In particular, if either one of the latter two conjectures holds, then there is no $2 - \epsilon$ approximations algorithm for the minimum maximal matching problem.

Comments:

- The results are technically solid although some of the reductions bear similarities to existing hardness results.
- The descriptions of the reductions are in a very compact form and thus not easy to follow.

Proportional Approval Voting System

The proportional approval voting system is an electoral system in which each voter chooses a committee of k members by checking the approval box next to the candidates. There is no limit on the number of approved candidates. In this electoral system, the score of a set of candidates is determined by a function $\mathrm{scr}_w()$, where w is a non-increasing sequence, and the winner is the set of k candidates whose score is maximum.

While the election is relatively simple, deciding the winner is NP-hard, and it has a connection to the well-studied multi-ℓ-coverage problem. Thus, the researchers lean forward to showing an approximate winner's score (although it would not be practical in actual elections).

Previously, the best approximation ratio was $(1 - 1/e)$ using the same greedy algorithm as the multi-ℓ-coverage problem. The authors improve over the long-standing bound by employing a linear program, yielding an $\alpha(w)$-approximation algorithm, where $\alpha(w)$ is a function that depends on the weight-sequence w. For all the studied weight functions, the ratio $\alpha(w)$ is better than $(1 - 1/e)$.

The authors also show that the factor-alpha(w) is the best possible unless P=NP. The authors also show that assuming Gap-ETH, the running time of alpha(w) must be at least $f(k) \cdot (|V| + |C|)^{o(k)}$. In addition, the authors present an instance where a greedy algorithm has a strictly worse approximation ratio than the LP-based algorithm presented in this thesis.

Comments:

- The positive result is solid, and the rounding algorithm, where the values are transformed into $\{0, q, 1\}$, is quite clever.
- The negative results mostly follow from the known result on the maximum k-coverage problem, but they are not trivial.

Capacitated k-Median

The capacitated k-median is a notorious problem. The only proper approximation algorithm to this problem employs a greedy set-cover algorithm, which gives a factor $O(\log k)$-approximation. Constant-factor approximation algorithms are known only when either the parameter k or the capacity is allowed to be violated by a factor $(1+\epsilon)$.

This thesis explores a different direction in relaxing this problem. While the violation of the number of centers and capacities is not allowed, the running time can be exponential on k. In particular, the authors present a $(7+\epsilon)$-approximation algorithm that runs in time $2^{O(k \log k)}\text{poly}(n)$, where n is the size of the instance. Thus, when the parameter k is a fixed constant, the running time is polynomial on the input size, thus giving one of the not-so-many results on FPT-approximation algorithms, which is becoming an active area in recent years.

The key technique is to reduce the input to an ℓ-centered instance, where we pick a set of l facilities as centers. The value of ℓ is chosen to be much larger than ℓ since we only need it to reduce the time complexity of the algorithm.

Next, the authors assign each client to the closest centers, thus forming a Voronoi diagram with l cells. This partitions the clients into disjoint subsets where clients in each group are in the same proximity. (Note that this method resembles the epsilon-net technique.)

Now, we can relay the demand to route through the center of each Voronoi cell. Henceforth, our task is to guess the number of facilities needed to open in each cell. There are at most ℓ^k possibilities. Thus, this gives a constant approximation algorithm that runs in $\ell^k\text{poly}(n)$ time. By choosing an appropriate value of ℓ, the authors give an $f(k)\text{poly}(n)(7+\epsilon)$-approximation algorithm for CkM.

Comments:

- The algorithmic result is very nice, and the direction for attacking this problem is quite natural.
- This chapter of the thesis has quite a few typos, while no typos are spotted in other chapters.

Typographical Error

- Page 52 Line 22: this problems $\Longrightarrow$ problem
- Page 53 Line 12: a S subset $\Longrightarrow$ a subset S
- Page 56 Line 2: a $O(\log k)$ $\Longrightarrow$ "an" Oh $\log k$
- Page 59 Line -6: The reason is, that $\Longrightarrow$ "is that" : remove comma
- Page 60 Line 10: Once we know, how $\Longrightarrow$ "know how" : remove comma

Other Comment

The Label-Cover problem defined in this paper is a variant that appeared in [Fei6] and in Moshkovitz's paper:

"The Projection Games Conjecture and the NP-Hardness of $\ln n$-Approximating Set-Cover"

The author may remark that this is not a standard version of the Label-Cover problem that is used in literature.