THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Department of Mathematics
London School of Economics and Political Science
Houghton Street
London, WC2A 2AE
UK

N.Olver@lse.ac.uk

May 16, 2025

Referee report for "Approximation Algorithms for Clustering and Submodular Facility Location Problems" by Fateme Abbasi

This dissertation provides a number of new results on the topic of approximation algorithms for clustering and related problems. The results concern three distinct problems.

- **Socially fair (k, z)-clustering.** This is a natural generalization of various well-studied clustering problems such as k-median, k-center and k-means. There are a number of groups, and the goal is to open k facilities such that the largest amongst all groups of the ℓ_z-norm of the distances of the clients in that group to their closest open facility is as small as possible. The problem had already received some attention in the past. The main contribution of this thesis to the topic is an improved FPT approximation for this problem in Euclidean metrics (of arbitrary dimension); that is, an approximation algorithm whose running time is of the form $f(k) \cdot \text{poly}(n)$, where n denotes the size of the metric space. A significant feature of the result is that it shows a separation between general metrics and Euclidean metrics: the approximation factor obtained is strictly better than what is possible for general metrics, under a standard complexity assumption.

  The algorithm is based heavily on the previous work for general metrics, but the analysis has a number of interesting and novel ideas in order to exploit the Euclidean structure.

  This result was published in ICALP, which is a very highly ranked conference in theoretical computer science.

- **An EPAS for clustering with a general norm.** In *norm k-clustering*, the goal is to minimize a specified norm f of the vector of distances between clients and their closest open center. This is a very general setting, especially if the norm is not required to be symmetric, i.e., invariant under permutations of the input vector coordinates. For example, it includes socially fair (k, z)-clustering as a special case.

  The main result of the thesis on this topic is an efficient parameterized approximation scheme (with parameter k), under some restrictions on the metric space. The metrics considered include well-studied ones such as continuous Euclidean metrics ("continuous" meaning that any point in Euclidean space is a potential location for a center), discrete Euclidean metrics of constant dimension, bounded treewidth metrics, and planar metrics.

  The result here is very technically sophisticated. It is quite impressive that such strong results can be obtained in this generality. This work was published in FOCS, which (along with STOC), is the top conference in theoretical computer science, and extremely competitive.

- **An improved approximation algorithm for submodular facility location.** In the submodular facility location problem, along with the metric space and set of clients and facility locations, a submodular function f on the set of facility locations is given. The goal is to open a set of facilities X such that the sum of distances of clients to closest facilities,

plus $f(X)$, is minimized.

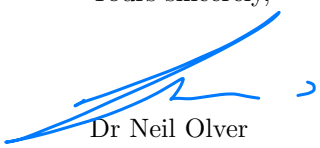
I have a particular fondness for this problem; I have thought about it, and related problems, over the years. The result of this thesis is an improved $O(\log \log n)$-factor approximation algorithm, where n is the size of the metric space. I was quite excited upon hearing of this result; until this point, the best we knew was an $O(\log n)$-factor approximation using very standard techniques.

The approach involves two main steps, both of which are interesting. The first step, although in hindsight not overly complicated, I find particularly nice. It shows that the problem can be reduced to an HST (a special kind of tree metric) losing only an $O(\log \log n)$ factor. I found this surprising, since naïvely, one would expect such a reduction to necessarily lose an $O(\log n)$ factor. The second step involves generalizing a result a student and I obtained for a related (but distinct) problem; essentially, showing that the approach works not only on a line, but also on an HST.

This result, like the first, was published in ICALP.

All three results are interesting and novel, and require significant conceptual and technical ideas. The writing is overall quite clear, and I believe all the results to be correct. In my estimation, these results make a clear contribution to the field, and are at the appropriate level for a PhD thesis. So I have no hesitation recommending that the process continue to the next stage, so that the candidate may defend her thesis.

Yours sincerely,



Dr Neil Olver