UNIVERSITY
OF WARSAW

Faculty of Mathematics, Informatics and Mechanics
Institute of Informatics

Warsaw, 15th April 2025

PhD Committee
University of Wrocław

Report on the PhD thesis of Fateme Abbasi
Approximation Algorithms for Clustering and Submodular Facility Location Problems

Background of the thesis. The presented research in the thesis belongs to the area of the theory of algorithms for clustering in metric spaces.

The family of k-*Clustering problems* is a wide umbrella for problems that ask for a specific number of clusters k (this is a hard constraint); a cluster is defined by choosing its center and then partitioning the set of points (sometimes called also clients) between centers. In most problems, a client can be greedily connected to the nearest center. The total cost of the solution is an aggregate function of the vector of all distances between clients and their respective centers; for example, the k-Center problem measures the maximum of the distances, k-Median the average, while more involved measures can include some fairness conditions.

Usually little can be done in polynomial time for this family of problems. However, it turns out that if one allows fixed-parameter algorithms with k as the parameter, we discover a very rich algorithmic landscape.

A closely related problem is Facility Location, where the number of centers is not strictly constrained to a fixed number k, but opening each facily incurs

Banacha 2, 02-097 Warsaw, Poland
+48 22 55 44 419
e-mail: m.pilipczuk@mimuw.edu.pl

some additional cost. In the vanilla version, each center has its opening cost (but after opening, it can serve any number of clients); in the thesis a more general version is studied, where the opening costs are governed by a submodular function of the set of clients attached to the center.

Main results. The PhD thesis of Fateme Abbasi is based on three conference papers, two from ICALP 2024 and one from FOCS 2023.

1. The first paper from ICALP 2024 tackles the k-Clustering problem, in a quite general form capturing e.g. all fairness conditions, in Euclidean spaces. If the norm involves z-powers of the distances, there is a known lower bound on the approximation factor of 3^z in general metric spaces (for fixed-parameter approximation algorithms, parameterized by k). The main result is that this barrier can be broken for Euclidean spaces (even for large dimensions).
   On the technical side, the starting point is the observation that the 3^z approximation factor appears in a very specific situation that can be avoided in Euclidean spaces by additionally looking at midpoints of segments spanned by an bicriteria-approximate solution. (The Euclidean assumption is essential to be able to define the midpoint of a segment here.) With this idea at hand, a highly technical and involved analysis is needed to prove that a slightly better approximation factor is possible.
2. The paper from FOCS 2023 is a landmark paper that introduces the notion of *ε-scatter dimension* that already gained significant interest in international community and resulted in a number of follow-up works. In short, the work introduces an invariant of the metric space — ε-scatter dimension — that turns out to govern the running time of parameterized approximation schemes for a very wide range of k-Clustering problems. This invariant turns out to be bounded in many classes of metric spaces (the paper shows it for metrics of bounded doubling dimension and those spanned by planar graphs; later works by other authors show that all excluded-minor metrics share this property). This gives a clean explanation for existence of efficient approximation algorithm in these classes of metrics.
   This part has a very strong and excellent theory-building component

Banacha 2, 02-097 Warsaw, Poland
+48 22 55 44 419
e-mail: m.pilipczuk@mimuw.edu.pl

that is already appreciated in the community. Even if the candidate did not directly contribute to the development of the new theory,[1] participation in such work is already a significant experience and, moreover, the work involves a wide selection of technical details and contributions, both in the algorithmic part and in the part that proves boundedness of the new invariant in classes of metrics.

3. The second paper from ICALP 2025 tackles the Submodular Facility Location problem, showing a $\mathcal{O}(\log \log n)$-approximation algorithm. I found the algorithm here particularly elegant, as it is mostly based on a clever combination of already known tools. First, we solve an LP relaxation of a configuration ILP for the problem, treat it as a probability distribution over pieces of solution, and sample from it $\mathcal{O}(\log \log n)$ times. This incurs acceptable cost, while (in expectation) decreasing the connection cost of an optimum solution by a $\mathcal{O}(\log n)$ factor. Being now able to overpay by a $\mathcal{O}(\log n)$ factor in the connection part of the cost, we use classic results on embeddings into a tree metric. The last stage is an observation that the resulting problem on a tree is similar to other problems that are already known to admit good approximation algorithms and one can adopt their approaches.
   This part is a nice play with a number of different tools from the literature, requiring somewhat wide horizons and knowledge. The final product is elegant and clean.

Evaluation. The presented results are very strong. The FOCS 2023 paper has a landmark theory-building component. All three presented works are involved and interesting from the technical point of view. Finally, all included papers tackle problems interesting for the international community. Overall, the scientific content of the thesis is far above the bar for a PhD thesis in computer science.

The thesis does not include any formal contribution statement of the candidate. From an informal inquiry I learned that the candidate was a major contributor to the first paper, had a fair contribution to the technical side of the second paper, and contributed significantly to preliminary research in the last work. Finally, she contributed a fair amount of writing up work

[1] The paper has many senior coauthors and no detailed contribution statement is available.

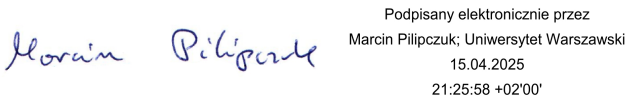
Banacha 2, 02-097 Warsaw, Poland
+48 22 55 44 419
e-mail: m.pilipczuk@mimuw.edu.pl

in all papers.

The first work is quite heavy on the technical side, hence being a major contributor here is an excellent achievement. The second work is very wide and massive, hence even a partial contribution to the technical side is substantial. Finally, the preliminary work done for the third paper, while not explicitly visible in the final product, helped guiding the research. Overall, the candidate contributed significantly to the presented results.

Summary. In my opinion the presented thesis satisfactorily meets all formal and customary requirements for a PhD thesis in computer science and I recommend admitting the candidate to the defense.

Yours faithfully,



Marcin Pilipczuk

Banacha 2, 02-097 Warsaw, Poland
+48 22 55 44 419
e-mail: m.pilipczuk@mimuw.edu.pl