Testing of Clustering, correction

Noga Alon Seannie Gilboa Michal Parnas Dana Ron

July 4, 2023

The k-diameter problem in the d-dimensional Euclidean space is the problem of partitioning a given set of n points in R^d into k clusters in a way that minimizes the maximum diameter of a cluster.

In Section 4.1, page 293 of [1] we remark, citing private communication from Leonard Schulman, that this problem can be solved exactly in $(O(n))^{dk^2}$ time.

Henry Fleischmann, Kyrylo Karlov, C. S. Karthik, Ashwin Padaki and Stepan Zharkov [2] pointed out that this claim is incorrect. Indeed, it is claimed in [1] that the problem can be reduced to considering only clusterings in which the convex hulls of the clusters are disjoint, since if a point of one cluster in an optimal clustering is in the convex hull of the points of another cluster it can be moved without increasing the diameter of either cluster. While this is true, this process does not reduce the problem to clusters with disjoint convex hulls. In addition, as pointed out by [2], it can be shown that the 3-diameter problem in Euclidean space is APX-hard, even in $O(\log n)$-dimensions. Therefore, the existence of the claimed algorithm would provide an $n^{O(\log n)}$ algorithm for an NP-hard problem, contradicting the Exponential Time Hypothesis.

While this remark is only marginally related to the main results of [1], it is sometimes interesting to analyze the running time of testing algorithms, and not only their query complexity.

Acknowledgments

We thank Henry Fleischmann, Kyrylo Karlov, C. S. Karthik, Ashwin Padaki, and Stepan Zharkov for their careful reading and insightful input.

References

[1] N. Alon, S. Dar, M. Parnas, and D. Ron, Testing of clustering, SIAM Review 46 (2004), 285-308.

[2] Henry Fleischmann, Kyrylo Karlov, C. S. Karthik, Ashwin Padaki, and Stepan Zharkov, Private Communication, 2023.