Approximability of Euclidean k-center and k-diameter

Todor Antic, Guillermo Gamboa Quintero, Jelena Glisic and Patrik Zavoral

REU24

Motivation

Data clustering: A way of dividing objects (data points) into groups (clusters) such that members of the same group are similar.

We might want to optimize for different parameters, such as:

- Average size of a cluster
- Maximal distance from the center of the cluster (k-center)
- Size of the biggest cluster (Max-k-diameter)

Definitions

- ***Def.*** Let (X, dist) be a metric space and $C \subset X$. Then

  $\mathrm{diam}(C) := \max \{ \mathrm{dist}(x,y) \mid x, y \in C \}$.

- ***Def.*** For a collection of subsets $C_1, C_2, \ldots, C_k \subset X$,

  $\mathrm{diam}(\{C_1, C_2, \ldots, C_k\}) := \max \{ \mathrm{dist}(x, y) \mid i \in [k] \ \& \ x, y \in C_i \}$.

Problem

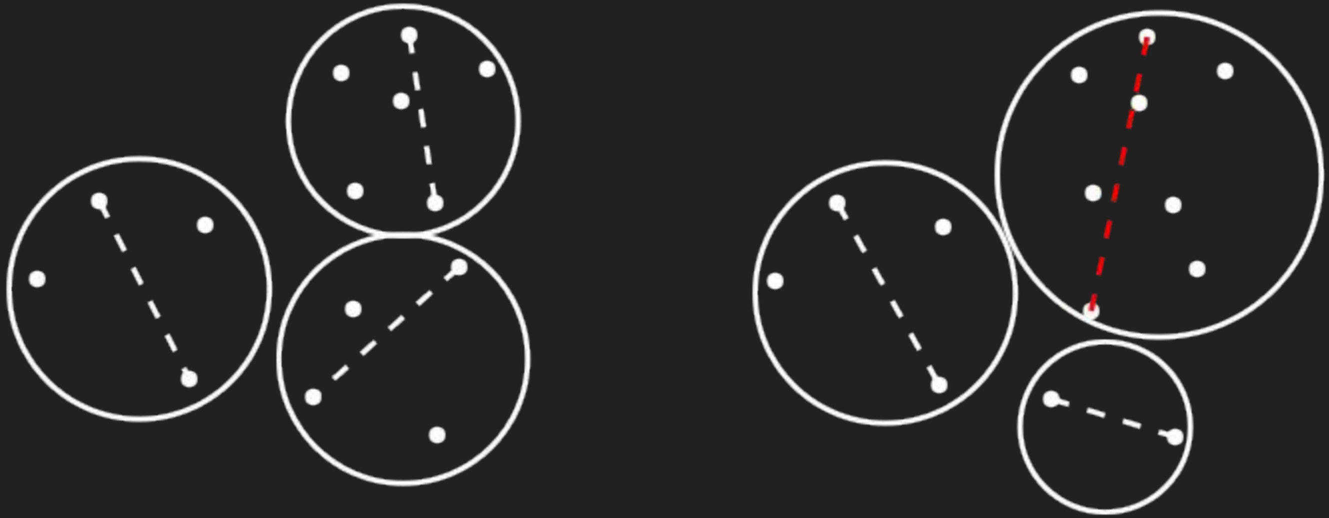
- **Max-k-Diameter** - let (X, dist) be a metric space and let k be a constant. Given as input $P \subset X$, find a k-clustering that minimizes $\mathrm{diam}(\{C_1, C_2, ..., C_k\})$.

- **r-approximate Max-k-Diameter** - given (X, dist), k, and input $P \subset X$, let

  $\Delta := \min \mathrm{diam}(\{C_1, C_2, ..., C_k\})$ over all possible clusterings of P.

  Find a k-clustering with diameter at most $r\Delta$.

Problem



State-of-the-art Approximability Results

Metric	NP-Hardness Approximation Factor	Polynomial Time Approximation Factor
ℓ_∞	**2-ε** [Meg90]	**2** [Gon85]
ℓ_0/ℓ_1	**1.5-ε** [FKSPZ24]	**2** [Gon85]
ℓ_2	**1.304** [FKSPZ24]	**1.415** [BHI02]

Goals of the Project

We want to:

- Study the problem in depth;
- Close the gaps in state-of-the-art.

Barriers to overcome:

- Intermediate distance in ℓ_1;
- Large odd girth in $\ell_{2.}$

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 823748.

References

[BHI02] Mihai Badoiu, Sariel Har-Peled, and Piotr Indyk. *Approximate clustering via core-sets*. In Proceedings of the thirty-fourth annual ACM symposium on Theory of computing,pages 250–257, 2002.

[FKSPZ24] Henry Fleischmann, Kyrylo Karlov, Karthik C. S., Ashwin Padaki, and Stepan Zharkov. *Inapproximability of Maximum Diameter Clustering for Few Clusters*. Arxiv preprint. 2024.

[Gon85] Teofilo F. Gonzalez. *Clustering to minimize the maximum intercluster distance*. Theoretical Computer Science, 38:293–306, 1985.

[Meg90] Nimrod Megiddo. *On the complexity of some geometric problems in unbounded dimension*. Journal of Symbolic Computation, 10(3–4):327–334, 1990.