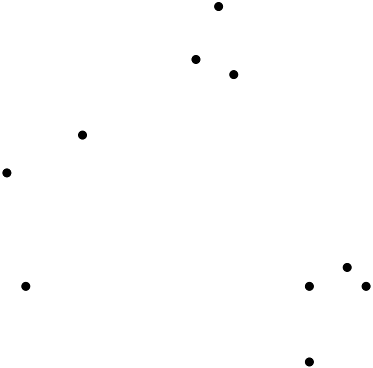
Hardness of Approximation for Clustering Objectives

Kyrylo Karlov   Ashwin Padaki   Styopa Zharkov

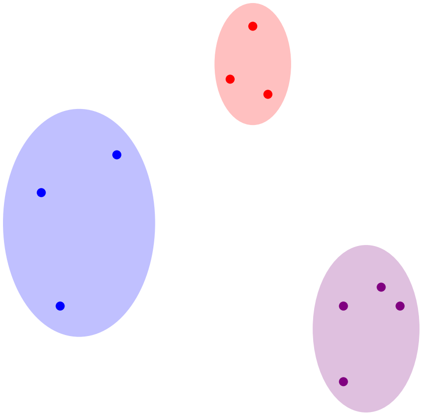
DIMACS REU 2023

DIMACS
Center for Discrete Mathematics and Theoretical Computer Science
Founded as a National Science Foundation Science and Technology Center

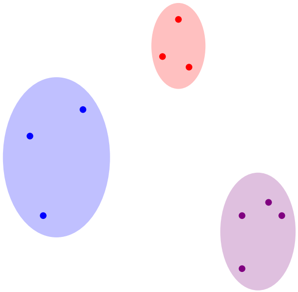
Clustering Problem



Clustering Problem



Clustering Problem

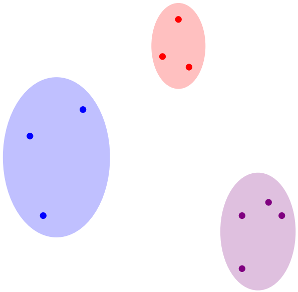


Definition

A k-clustering of a set of points is a partition $\mathcal{P}$ of the points into k sets.

DIMACS
Center for Discrete Mathematics and Theoretical Computer Science
Founded as a National Science Foundation Science and Technology Center

Clustering Problem

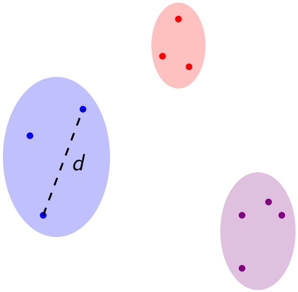


Definition

A k-clustering of a set of points is a partition $\mathcal{P}$ of the points into k sets.

Problem: Find a k-clustering that optimizes some objective function.

Objective Function: Diameter



A 3-clustering with (Euclidean) diameter d

Definition

The **diameter** of a clustering is the maximum distance between two points in the same cluster.

Computational Complexity of Clustering

Clustering is a computationally hard problem!

- Finding the k-clustering that minimizes diameter is **NP-Hard**.

Computational Complexity of Clustering

Clustering is a computationally hard problem!

- Finding the k-clustering that minimizes diameter is **NP-Hard**.
- We can consider approximate solutions instead.

DIMACS
Center for Discrete Mathematics and Theoretical Computer Science
Founded as a National Science Foundation Science and Technology Center

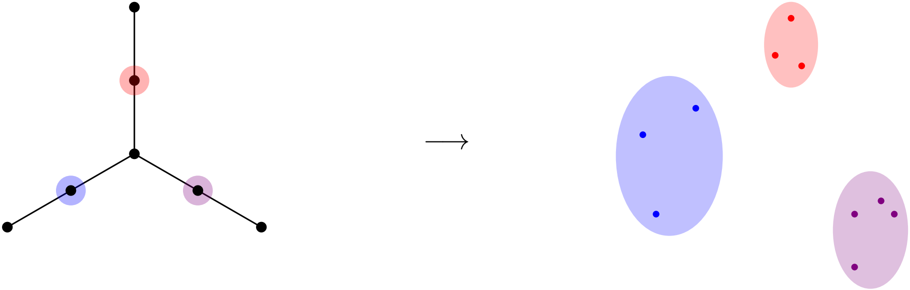
Hardness depends on the **approximation factor** and **objective function**.

- There is an efficient 2-approximation algorithm for diameter.
- In many cases, we can show that any better approximation is hard.

Hardness in the ℓ_1 Metric

Example

Better than 2-approximation for diameter with ℓ_1 metric is hard.



Hardness in the ℓ_2 Metric

Example

≈ 1.97-approximation for diameter with ℓ_2 metric is hard.

Question: Is efficient 1.98-approximation possible?

Research Goal: Close these gaps!

Acknowledgements

- **Mentors:** Karthik C.S., Henry Fleischmann
- This work will be carried out while Ashwin Padaki and Styopa Zharkov are in the 2023 DIMACS REU program at Rutgers University, supported by NSF grant CNS-2150186.
- This work will be carried out while Kyrylo Karlov is a participant in the 2023 DIMACS REU program at Rutgers University, supported by CoSP, a project funded by European Union's Horizon 2020 research and innovation programme, grant agreement No. 823748.