

## K-medoids

Given samples  $\mathcal{X} = \{x(1), \dots, x(N)\}$ , function  $dist : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}^+$  and monotonic energy function  $\psi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ , find  $\mathcal{C} \subset \{1, \dots, N\}$  where  $|\mathcal{C}| = K$  to minimize

$$E = \sum_{i=1}^N \min_{i' \in \mathcal{C}} \psi(dist(x(i), x(i'))).$$

Applications for clustering sequences, graph vertices, sparse and dense vectors, etc. Two popular algorithms are,













- MEDLLOYD (Hastie et al. (2001), Park and Jun (2009)): like Lloyd's algorithm, but centroids are replaced by medoids
- CLARANS (Ng and Han (1994)): random swaps between centers and non-centers are proposed, and only accepted if  $E(\mathcal{C})$  decreases.

## K-means and K-means seeding

The K-means task is to find  $K$  centers,  $\{c(1), \dots, c(K)\}$ , *not necessarily elements of*  $\{x(1), \dots, x(N)\}$ , to minimize

$$E = \sum_{i=1}^N \min_{k \in \{1, \dots, K\}} \|x(i) - c(k)\|_2^2. \quad (1)$$

Centers for the popular Lloyd's algorithm (a.k.a. the K-means algorithm) are commonly initialized or *seeded* as a subset of  $\mathcal{X}$ . Good seeding is critical for to avoid poor local minima.

(A)	seed:	  	$E = 4$
	final:	  	$E = 2$
(B)	seed:	  	$E = 3.38$
	final:	  	$E = 1.69$

Case (B) has lower seeding *and* final  $E$ . Most seeding algorithms attempt to minimize initial energy (k-means++, Bradley-Fayad, etc.). Minimizing seeding energy is the special case of K-medoids with

$$dist(x(i), x(i')) = \|x(i) - x(i')\|_2 \quad \text{and} \quad \psi(v) = v^2,$$

which motivates the use of popular and established K-medoids algorithms for K-means seeding.

## Acknowledgements

This work was sponsored by **HASLERSTIFTUNG**

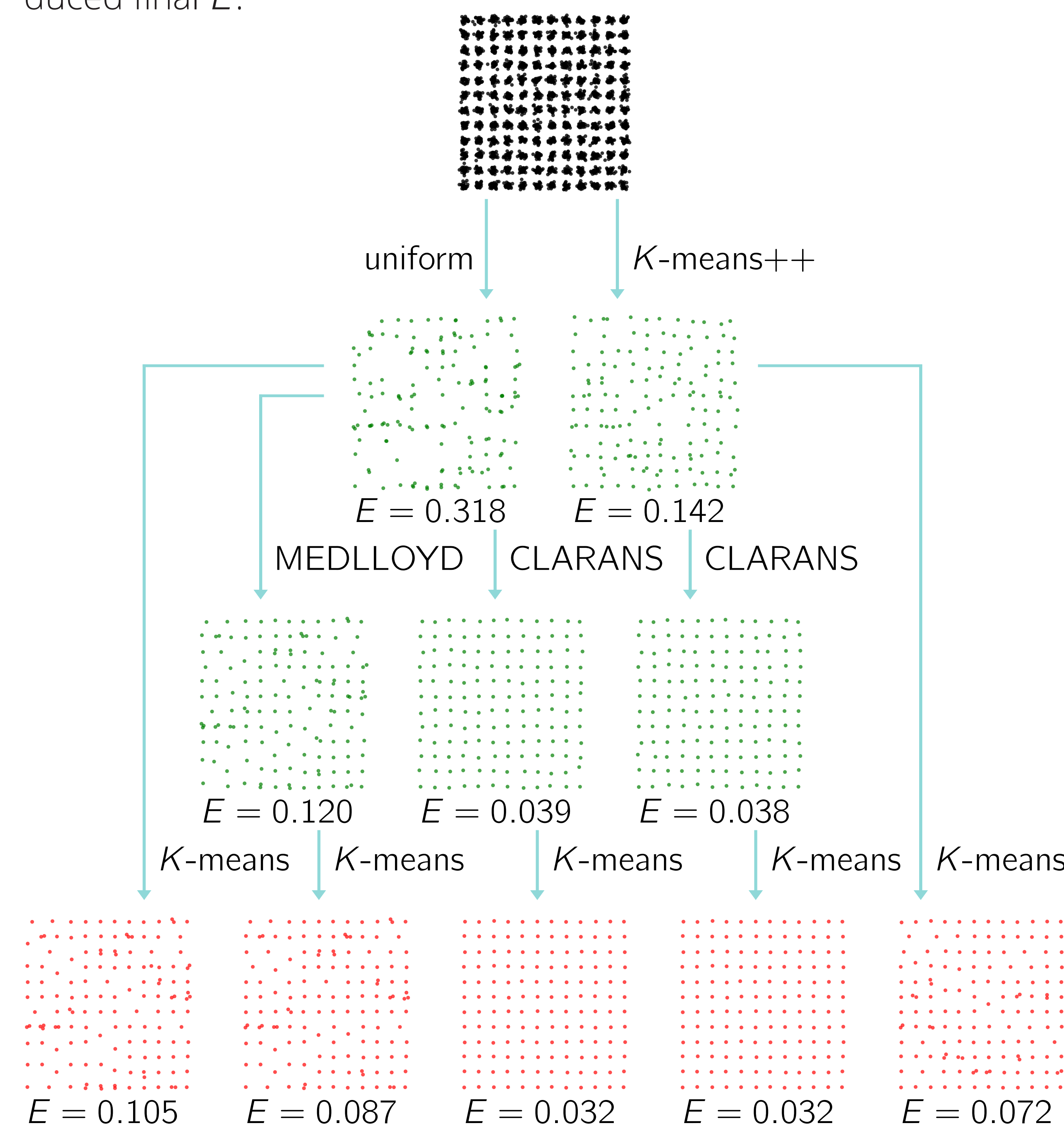
## The CLARANS K-medoids algorithm

The algorithm iteratively proposed swapping a medoid ( $x(i_-)$ ) with a non-medoid ( $x(i_+)$ ). Energy reducing swaps are implemented.

- 1: Initialize center indices  $\mathcal{C} \subset \{1, \dots, N\}$ , where  $|\mathcal{C}| = K$
- 2:  $E \leftarrow \sum_{i=1}^N \min_{i' \in \mathcal{C}} \psi(dist(x(i), x(i')))$
- 3: **while** stopping criterion false **do**
- 4:   sample  $i_- \in \mathcal{C}$  and  $i_+ \in \{1, \dots, N\} \setminus \mathcal{C}$
- 5:    $E^+ \leftarrow \sum_{i=1}^N \min_{i' \in \mathcal{C} \setminus \{i_-\} \cup \{i_+\}} \psi(dist(x(i), x(i')))$
- 6:   **if**  $E^+ < E$  **then**
- 7:      $\mathcal{C} \leftarrow \mathcal{C} \setminus \{i_-\} \cup \{i_+\}$
- 8:      $E \leftarrow E^+$
- 9:   **end if**
- 10: **end while**

## Five routes to K-means local minima

(Below) Clustering with  $K = 12^2$  centers on a 2-d grid,  $K = N/25$ . First row, the generated samples. Second row, uniform and K-means++ seedings. Third row, K-medoids refinements. Fourth row, final K-means clusterings. CLARANS refinement results in reduced final  $E$ .



## Accelerating the CLARANS algorithm

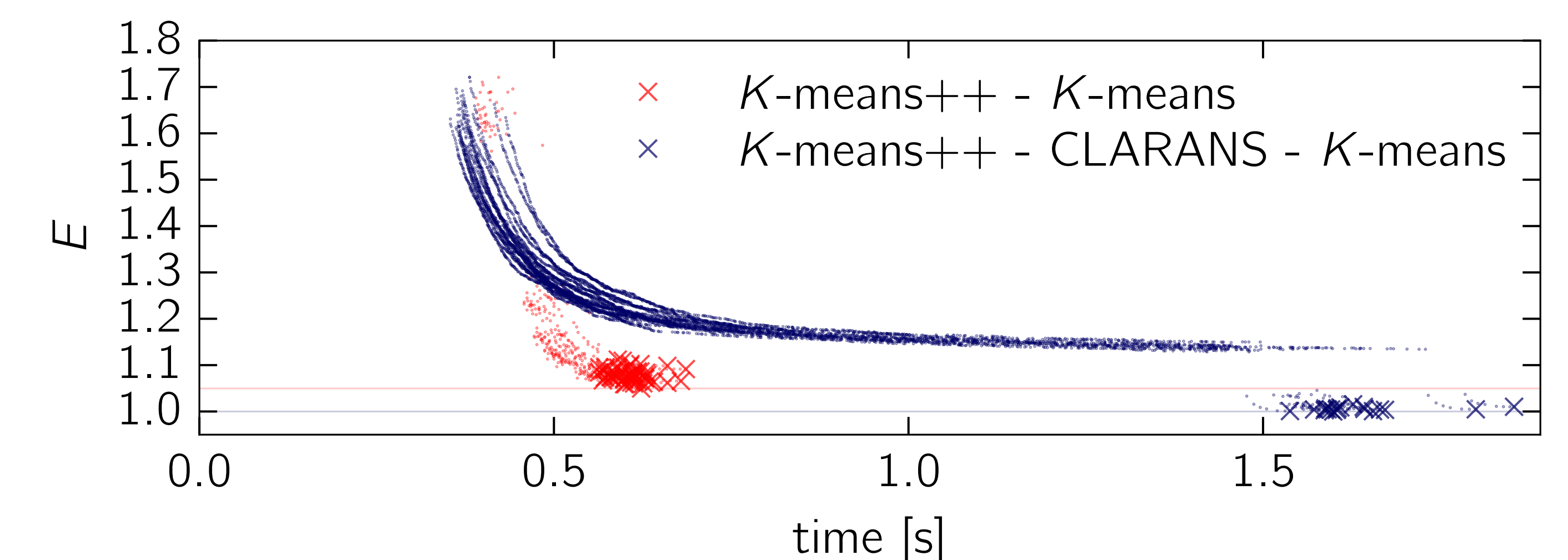
There are many more *evaluations* (line 5) than actual *implementations*. Assuming balanced clusters, unoptimised evaluation complexity is  $O(N)$ . When  $dist$  satisfies the triangle inequality, we present a technique with evaluation complexity  $O(K)$ , implementation complexity  $O(N)$ . It requires recording,

- for non-centers (such as  $x(1)$  below), distance to nearest ( $d_1$ ) and second nearest ( $d_2$ ) centers
- for centers (such as  $x(2)$ ), maximum over cluster of  $d_1$  and  $d_2$  ( $R_1$  and  $R_2$  respectively), and distances to all centers.

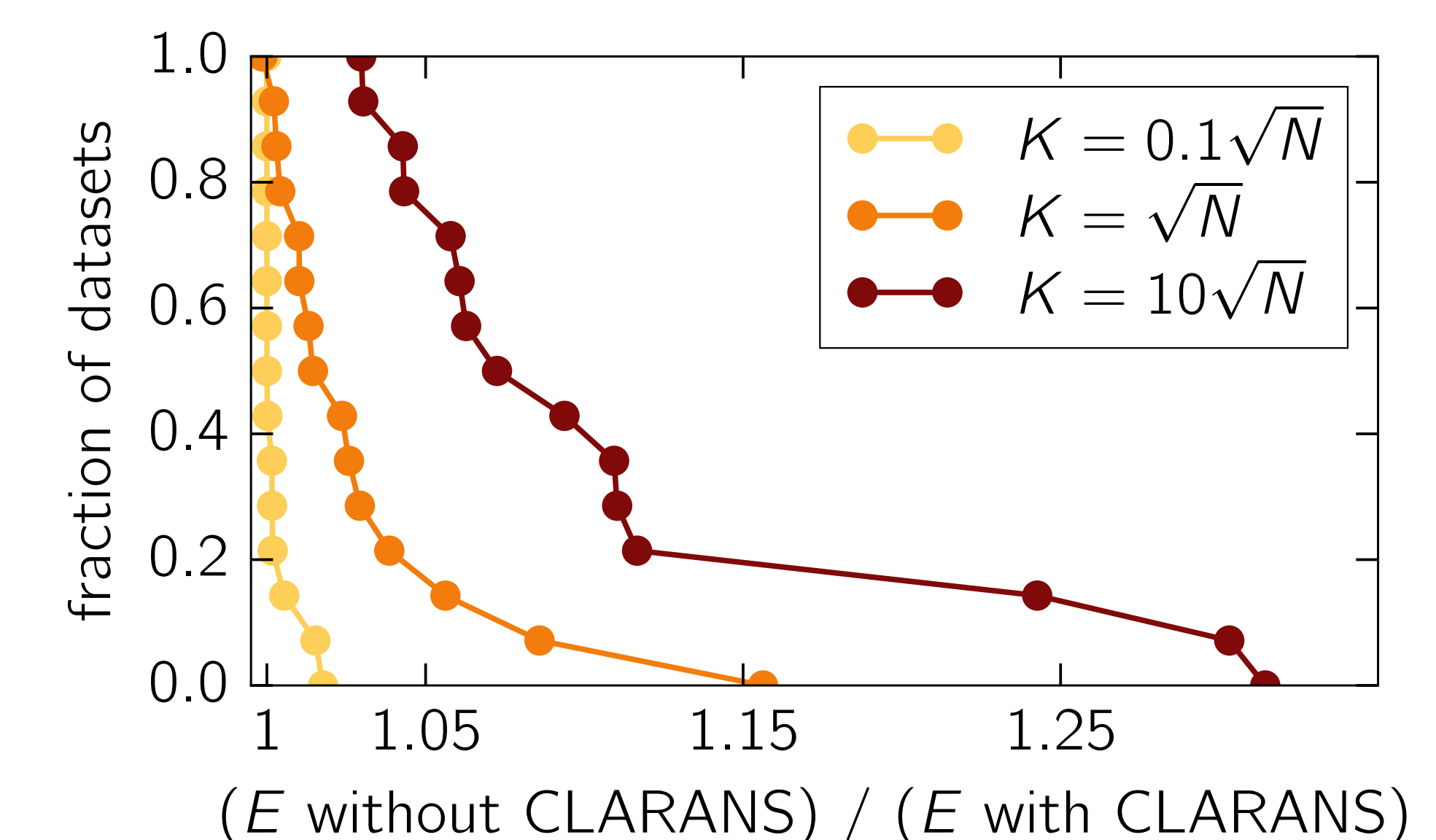


## Results

(Below) An experiment with an RNA dataset,  $N = 16 \times 10^4$ ,  $d = 8$  and  $K = 4 \times 10^2$ . With 50 runs seeded with K-means++ (red), and several runs with CLARANS between K-means and K-means++ (blue). The best run without CLARANS has 5% higher  $E$ .



(Below) Summary of experiments on 16 publicly available datasets. Each point is an experiment with same setup as for RNA above, horizontal position is reduction in  $E$  obtained using CLARANS. Datasets have dimensions  $d : 2 \rightarrow 90$ ,  $N : 1484 \rightarrow 488565$ .



code on github