Local Search for Clustering in Almost-linear Time

Abstract

We propose the first *local search* algorithm for Euclidean clustering that attains an $O(1)$-approximation in almost-linear time. Specifically, for Euclidean k-MEANS, our algorithm achieves an $O(c)$-approximation in $\widetilde{O}(n^{1+1/c})$ time, for any constant $c \geq 1$, maintaining the same running time as the previous (non-local-search-based) approach [la Tour and Saulpic, arXiv'2407.11217] while improving the approximation factor from $O(c^6)$ to $O(c)$. The algorithm generalizes to any metric space with sparse spanners, delivering efficient constant approximation in ℓ_p metrics, doubling metrics, Jaccard metrics, etc.

This generality derives from our main technical contribution: a local search algorithm on general graphs that obtains an $O(1)$-approximation in almost-linear time. We establish this through a new 1-swap local search framework featuring a novel swap selection rule. At a high level, this rule "scores" every possible swap, based on both its modification to the clustering and its improvement to the clustering objective, and then selects those high-scoring swaps. To implement this, we design a new data structure for maintaining approximate nearest neighbors with amortized guarantees tailored to our framework.