Fully Dynamic k-Center Clustering on graphs

Name*
Institution
email

ABSTRACT

Static and dynamic clustering algorithms are a fundamental tool in any machine learning library. Most of the efforts in developing dynamic machine learning and data mining algorithms have been focusing on the sliding window model (where at any given point in time only the most recent data items are retained) or more simplistic models. However, in many real-world applications one might need to deal with arbitrary deletions and insertions. For example, one might need to remove data items that are not necessarily the oldest ones, because they have been flagged as containing inappropriate content or due to privacy concerns. Clustering trajectory data might also require to deal with more general update operations.

We develop a $(2 + \epsilon)$-approximation algorithm for the k-center clustering problem with "small" amortized cost under the fully dynamic adversarial model. In such a model, points can be added or removed arbitrarily, provided that the adversary does not have access to the random choices of our algorithm. The amortized cost of our algorithm is polylogarithmic when the ratio between the maximum and minimum distance between any two points in input is bounded by a polynomial, while k and ϵ are constant. Our theoretical results are complemented with an extensive experimental evaluation on dynamic data from Twitter, Flickr, as well as trajectory data, demonstrating the effectiveness of our approach.

ACM Reference format:
Name. 2016. Fully Dynamic k-Center Clustering on graphs. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 1 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 RELATED WORK

In their article from 2004, Demetrescu and Italiano present an algorithm for the all pairs shortest paths problem. It works with a general graph with non-negative real-valued edge weights and supports any sequence of operations for an amortized running time of $O(n^2 log^3 n)$ per update.

An algorithm maintaining a distance matrix will work in $\Omega(n^2)$ as for each update, at most n^2 paths are changed. The above algorithm is thus essentially optimal.

In 2017, Abraham, Chechik and Krinninger came up with an algorithm taking a directed weighted graph with no unweighted cycles as input, against an adaptive online adversary, working in $O(n^{2+\frac{2}{3}} log^{\frac{4}{3}} n)$.

Other results for specific classes of graphs exist. In 1995, Henzinger, Klein, Rao and Subramanian found a data structure handling planar graphs in $O(n^{\frac{9}{7}} logD)$ per operation plus a preprocessing time of $O(n^{\frac{10}{7}})$. An operation being a query, an edge deletion an adge addition or chanching lenths. D being an upper bound on the sum of the absolute-values of the negative edge-lengths.

*Note

Conference'17, Washington, DC, USA

DOI: 10.1145/nnnnnnn.nnnnnnn