



Coresets for Clustering on Graphs with Bounded Treewidth



Dr. Shaofeng Jiang
Postdoctoral Researcher
Weizmann Institute of Science

🕒 2019年7月8日 星期一 10:00

📍 北京大学静园五院102



Abstract

We study coresets for clustering on graphs. Clustering is one of the central tasks in analyzing graph data. For instance, clustering is used for community detection on social networks, for data mining of road networks, and for visualization of large graph data sets. However, efficiently clustering on huge graph data sets remains a challenge. To this end, a power data reduction technique called coreset is introduced. A coreset is a compact summary of a large data set, such that the clustering objective evaluated on the coreset is preserved for all potential centers.

We obtain an efficient coreset construction for k -Median clustering on graphs. The size of our coreset is independent of the size of the data set, and it has only a linear dependence in treewidth which is a widely used complexity measure of graphs. We complement our coreset construction with a nearly matching lower bound, in which we show the linear dependence in treewidth is necessary for k -Median on graphs.

This talk is based on a joint work with Vladimir Braverman, Lingxiao Huang, Robert Krauthgamer and Xuan Wu.

Biography

Dr. Shaofeng Jiang is currently a postdoctoral researcher in the Weizmann Institute of Science, hosted by Robert Krauthgamer. Before joining Weizmann, he obtained his Ph.D. degree from the University of Hong Kong at 2017. His research interest is generally in theoretical computer science, and specially in algorithms for massive data sets, approximation algorithms and online algorithms. His works have been published in top venues including FOCS, SODA, ICML and INFOCOM. He is a recipient of MSRA Fellowship Nomination Award, and Outstanding Achievements in Postdoctoral Research Award in the Weizmann Institute of Science.