



A Coreset Framework for Constrained k-Clustering



Dr. Xuan Wu

TCS Lab
Huawei

🗣️ Host: 姜少峰 助理教授

🕒 2023年2月28日 星期二 16:00-17:00

📍 静园五院204室



Abstract

k-Clustering (e.g k-means, k-median) is the task to partition a data set into k groups of similar points. In practice, specific constraints are often imposed on the clustering task. Examples include capacitated clustering, fair clustering, robust clustering, and fault-tolerant clustering. Despite their importance, these constrained variants of clustering often do not admit a scalable solution or are even hard to approximate. To this end, we consider a data reduction technique, called coresets. An epsilon-coreset is a small subset of the input dataset such that the clustering objectives are almost the same (up to a $1+\epsilon$ factor). In the vanilla case, coresets are a justified scalable solution for k-clustering and have been extensively studied. However, in the constrained case, our understanding of coresets is falling way behind for a long while.

In this talk, I will introduce a new coreset framework that allows us to construct improved or new coresets for a large family of constrained clustering, based on a series of my recent works. In particular, I will show how to use the framework to construct the first $\text{poly}(k/\epsilon)$ -sized epsilon-coresets for capacitated, fair, and fault-tolerant k-clustering and the first coreset of size $O(m+\text{poly}(k/\epsilon))$ for (k,m) -robust clustering (m outliers should be excluded), in Euclidean space and other important metric spaces. Moreover, I will introduce a general approach to studying structural constraints of clustering problems and try to explore the limit of this framework.

Biography

Xuan Wu is a researcher in Huawei TCS Lab. Before joining Huawei, he obtained his Ph.D. from Johns Hopkins University. Prior to JHU, Xuan Wu earned his Master's Degree and Bachelor's Degree from IIS and Yao Class, both at Tsinghua University.