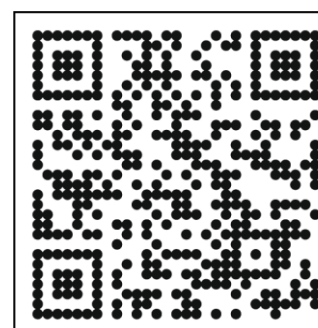




On Tight Robust Coresets for κ -Median Clustering



Dr. Xuan Wu

Nanyang Technological University

🎤 Host: 姜少峰 助理教授

🕒 2025年10月27日 星期一 15:00

📍 静园五院204室



Abstract

Designing efficient algorithms for robust clustering presents a significant challenge in algorithmic machine learning. Recently, a powerful data reduction technique called coreset has been applied to this problem, leading to active research in this direction.

In this talk, I will recap recent progress in coresets for robust clustering and introduce new coresets for the robust κ -median problem with m outliers, where new constructions in various metric spaces are obtained. Specifically, for metric spaces with a bounded VC or doubling dimension d , the coreset size is $O(m) + \tilde{O}(kd\epsilon^{-2})$, which is optimal up to logarithmic factors. For Euclidean spaces, the coreset size is $O(m\epsilon^{-1}) + \tilde{O}(\min\{k^{\frac{4}{3}}\epsilon^{-2}, k\epsilon^{-3}\})$ improving upon a recent result by Jiang and Lou (ICALP 2025). These results also extend to robust (k, z) clustering, yielding, for VC and doubling dimension, a coreset size of $O(m) + \tilde{O}(kd\epsilon^{-2z})$ with the optimal linear dependence on m . This extended result improves upon the earlier work of Huang et al. (SODA 2025). The techniques introduce novel dataset decompositions, enabling chaining arguments to be applied jointly across multiple components.

Biography

Xuan Wu is currently a research fellow in Nanyang Technological University. Prior to NTU, he has worked as a researcher in Huawei TCS Lab. He earned his Ph.D. from Johns Hopkins University and received his Master's Degree and Bachelor's Degree from Tsinghua University.