



# Metric Clustering and MST with Strong and Weak Distance Oracles



Dr. Chen Wang

Rice University & Texas A&M University

📍 Host: 孔雨晴 助理教授

🕒 2024年6月4日 星期二 10:30

📍 静园五院204室



## Abstract

I will discuss recent results of  $k$ -clustering and MST in a new weak-strong oracle model. In this model, for a fixed metric space  $(X, d)$ , we can compute distances in two ways: via a ‘strong’ oracle that returns exact distances  $d(x, y)$ , and a ‘weak’ oracle that returns distances  $\tilde{d}(x, y)$  which may be arbitrarily corrupted with some probability. This model captures the increasingly common trade-off between employing both an expensive similarity model (e.g. a large-scale embedding model) and a less accurate but cheaper model. Hence, the goal is to make as few queries to the strong oracle as possible. We consider both ‘point queries’, where the strong oracle is queried on a set of points  $S \subset X$  and returns  $d(x, y)$  for all  $x, y \in S$ , and ‘edge queries’ where it is queried for individual distances  $d(x, y)$ .

Our main contributions are optimal algorithms and lower bounds for clustering and Minimum Spanning Tree (MST) in this model. For  $k$ -centers,  $k$ -median, and  $k$ -means, we give constant factor approximation algorithms with only  $O(\tilde{k})$  strong oracle point queries, and prove that  $\Omega(k)$  queries are required for any bounded approximation. For edge queries, our upper and lower bounds are both  $\tilde{O}(k^2)$ . Surprisingly, for the MST problem, we give an  $O(\sqrt{\log n})$  approximation algorithm using no strong oracle queries at all, and we prove a matching  $\Omega(\sqrt{\log n})$  lower bound which holds even if  $\tilde{\Omega}(n)$  strong oracle point queries are allowed.

Based on a joint work with MohammadHossein Bateni, Prathamesh Dharangutte, and Rajesh Jayaram on COLT 2024.

## Biography

Chen Wang is a postdoctoral researcher hosted jointly by Vladimir (Vova) Braverman at Rice University and Samson Zhou at Texas A&M University. His research interests mainly focus on the intersections between Theoretical Computer Science and Machine Learning. In particular, he is interested in the theoretical foundations of practical learning problems and the design of algorithms with rigorous guarantees therein. More broadly, he is also interested in streaming algorithms and lower bounds, graph algorithms, statistical learning theory, and data processing privacy.

Chen obtained his Ph.D. at Rutgers University, where he was advised by Sepehr Assadi. He is the recipient of the Rutgers SGS Research & Travel Award, and he obtained nominations from Rutgers to Google Ph.D. fellow and Apple ML/AI scholar. He also received travel awards from various conferences, including SODA 2022, Neurips 2022, Neurips 2023, and STOC 2023.