Cluster: How hard is it to classify data?

Abstract:

Two popular objectives optimized in clustering algorithms are k-means and k-median. The k-means (resp. k-median) problem in the $L_p$-metric is specified by $n$ points as input and the goal is to classify the input point-set into $k$ clusters such that the k-means (resp. k-median) objective is minimized. The best-known inapproximability factor in literature for these NP-hard problems in $L_p$-metrics were well-below 1.01. In this talk, we take a significant step to improve the hardness of approximating these problems in various $L_p$-metrics.