

# Parallel Clustering Algorithms

Gyanesh Gupta Mengyang Zhang

NYU

## Project Plan

In this project, we aim to implement and study parallel clustering techniques used in data analysis. Specifically, we will be focusing on KMeans clustering. Within KMeans, we will extend the algorithm by using various initialization methods (KMeans++/Random partition), various variants (batch KMeans, KMedians/Medoids) and if possible, various methods (heirarchical clustering) as time permits. We will use various parallel programming methods discussed in the course (MPI/openmpi) and run on various datasets available. Finally, we will compare our code with various other implementations.