

# A fast hybrid clustering technique based on local nearest neighbor using minimum spanning tree

Gaurav Mishra, Sraban Kumar Mohanty, Tirumalesh Reddy Neelapu

October 23, 2022

## 1 The main idea

Most of the existing clustering algorithms become ineffective when inappropriate parameters are provided or applied on a dataset which consists of clusters of diverse shapes, sizes, and varying densities. To overcome such issues, graph-based hybrid clustering algorithms have been proposed. Researchers proposed a fast hybrid clustering technique based on local nearest neighbor using minimum spanning tree to reduce the computational overhead. Moreover, this algorithm does not require any user defined parameters and it can estimate the number of clusters more accurately.

## 2 The methodology

In this case study, reasearchers employed the divide-and-conquer technique to identify the actual clusters with improved efficiency.

1. Divide Step: For the given dataset of  $N$  data points, the dispersion level of data points is used to partition the dataset into sub-clusters.
2. Conquer Step: Minimum Spanning tree algorithm like Prim's or Kruskals can be used to construct the MST for each of the sub-clusters obtained from the Divide Step.
3. Combine Step: The Obtained sub-clusters are merged into actual clusters based on cohesion and intra similarity.

## 3 The results

Proposed algorithm is tested on 7 different datasets with number of clusters ranging from 2 to 8. And compared against other clustering algorithms like K-means, Hierarchical, DBSCAN, DMST, SAM and outperforms with all datasets. Also, the execution times were plotted comparing all the clustering algorithms and the proposed algorithm with MST proved to be executing in the least time comparing to every other algorithm.

## 4 Recommendation

Data preprocessing techniques were not used before applying the clustering algorithms. Removing the extreme outliers and other preprocessing techniques helps the model to cluster more accurately.