

Parallel computing is rapidly becoming an indispensable tool with the growing need for efficient processing and analyzing of data in all research disciplines. One particular algorithm that we are interested in is DBSCAN, a density-based clustering algorithm that classifies each data point to a cluster or noise, and the number of cluster is not known beforehand. It depends on two parameters, the distance that defines the neighbors of every data point, and the minimum number of neighbors in this neighborhood.

Parallelized DBSCAN has an immediate application to immunity researchers who wish to differentiate up to 10^7 cells based on 8 (cellular features) parameters.

Essential Features

- parallelization implemented on GPU
- weak scaling up to data size > 10 Mb

Extra Features

- neighbor seeking in the region query algorithm in $O(n \log n)$
- visualization

VISUALIZATION

