

For my parallelization of k-medoids, I first broke the process into two main steps:

- Assigning points to clusters
- Calculating new medoids

I then parallelized the independent, load balanced components of each of these steps, the distance calculation loops

For assigning points to clusters, I parallelized finding the distance from each point to the closest medoid, since the distance of each point to each medoid is independent of all other distances. Each thread independently tracks their own minimum distance, and can assign the point to a medoid without worrying about interference from other threads, and each thread has about the same amount of work to do.

For calculating the new medoids, I parallelized the computation of average distances for each point in a cluster. Each thread calculates the average distance for some point and then writes the result to an array. One final pass over this array is done to find the point with the lowest average distance, and that point is assigned as the medoid. I chose to parallelize at this level and not the cluster level because clusters will have very different numbers of points, there would be a load imbalance across threads. Within each cluster, each thread will need to do the same number of computations for each point, and load is more evenly distributed.

Below are the times I got, all executables were compiled with the -O3 optimization flag. Medoid calculation seemed to be the scaling issue for my Pthreads implementation for the 1024 and 512 group runs, either from overhead of waiting for threads to exit or some other error on my end, but to me it looks to follow pretty much structure as my OMP implementation.

PThreads 1024 groups

1 thread: 2833.1716s
2 threads: 1508.9991s
4 threads: 774.2665s
8 threads: 426.6087s
16 threads: 280.4274s

OMP 1024 groups

1 thread: 2820.0516s
2 threads: 1493.8217s
4 threads: 750.8355s
8 threads: 378.8152s
16 threads: 191.2793s

PThreads 512 groups

1 thread: 2575.5010s
2 threads: 1360.2163s
4 threads: 690.7266s
8 threads: 367.9381s
16 threads: 226.4813s

OMP 512 groups

1 thread: 2568.8325s
2 threads: 1352.2975s
4 threads: 676.5390s
8 threads: 342.0846s
16 threads: 178.3071s

PThreads 256 groups

1 thread: 8918.8467s
2 threads: 4508.7978s
4 threads: 2212.7195s
8 threads: 1104.8245s
16 threads: 571.5824s

OMP 256 groups

1 thread: 8914.5266s
2 threads: 4463.8635s
4 threads: 2238.7004s
8 threads: 1122.5413s
16 threads: 568.3704s