K-MEANS INITIALIZATION PROBLEM

Initial by center or partition

Zhiyuan Wang
12032878

SUSTech Southern University of Science and Technology

CONTENT

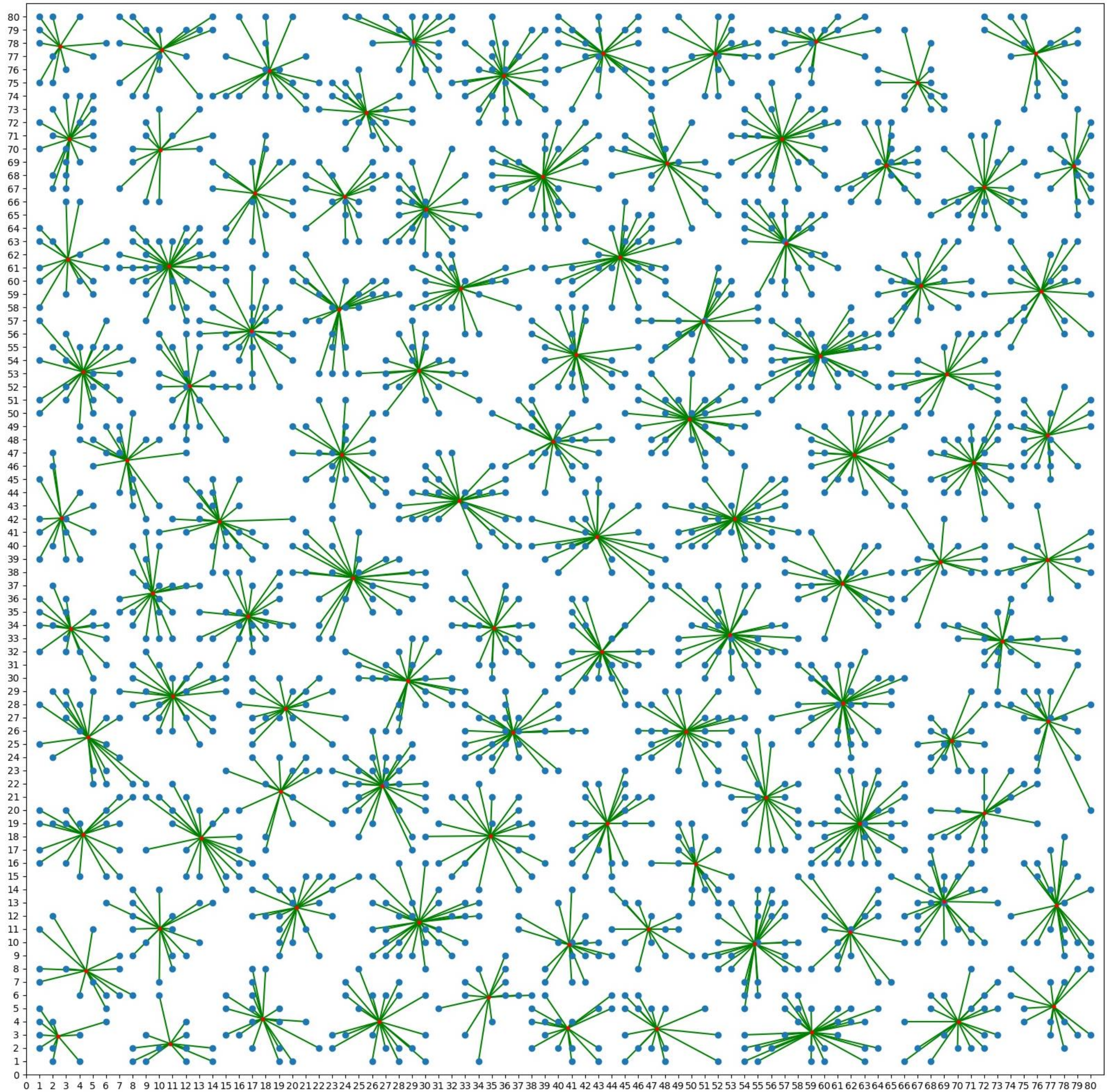
1. K-Means Algorithm
2. Initial K-Means by centers or partition
3. Initialization algorithm for K-Means

K-MEANS ALGORITHM

An algorithm to solve the clustering problem. It will iterate the following two steps from a random partition of S into k subsets: S_1, S_2, ..., S_k.

(i) $c_j = \frac{1}{|S_j|} \sum_{s \in S_j} s$

(ii) $S_j = \{s \mid \mathrm{dist}(s, c_j) = \min_{l=1,2,...,k} \mathrm{dist}(s, c_l)\}, j = 1, 2, \ldots, k$



INITIAL K-MEANS

- Initial by partition :
  - Generate K partition
  - Calculate the average coordinate as the center of partition.

- Initial by centers
  - Choice k site as the initial centers
  - Each site choice the nearest center as the center and join it’s partition.

INITIALIZATION ALGORITHM FOR K-MEANS

1. Initial by center:
   I. Use the center select algorithm to choice the initial centers.
   II. For each site s_i, select the first $\frac{1}{3}$ closest sites and let the maximum value of the distance be d_i, I will choice the site s_i which can minimize the d_i as the start center.
   III. Then, choice the farthest site from the centers as the next center iteratively until we get enough centers.

Algorithm 5 Greedy center selection algorithm

Require: S : all sites, k : The number of the center
Ensure:

```
function greedy_center_selection_algorithm(S, k)
    C← []
    C.add(S.index(random_int()))
    while C.size<k do
        C.append(argmax_s Distance(s, C))
    end while
    return C
end function
```

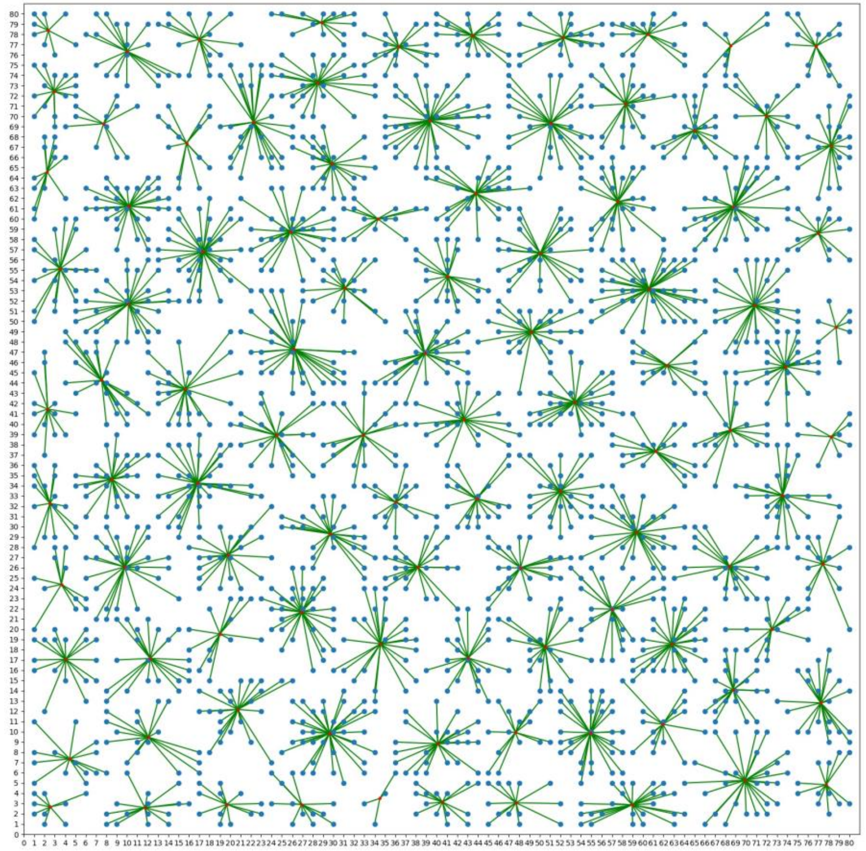
Algorithm 6 Initial center selection algorithm

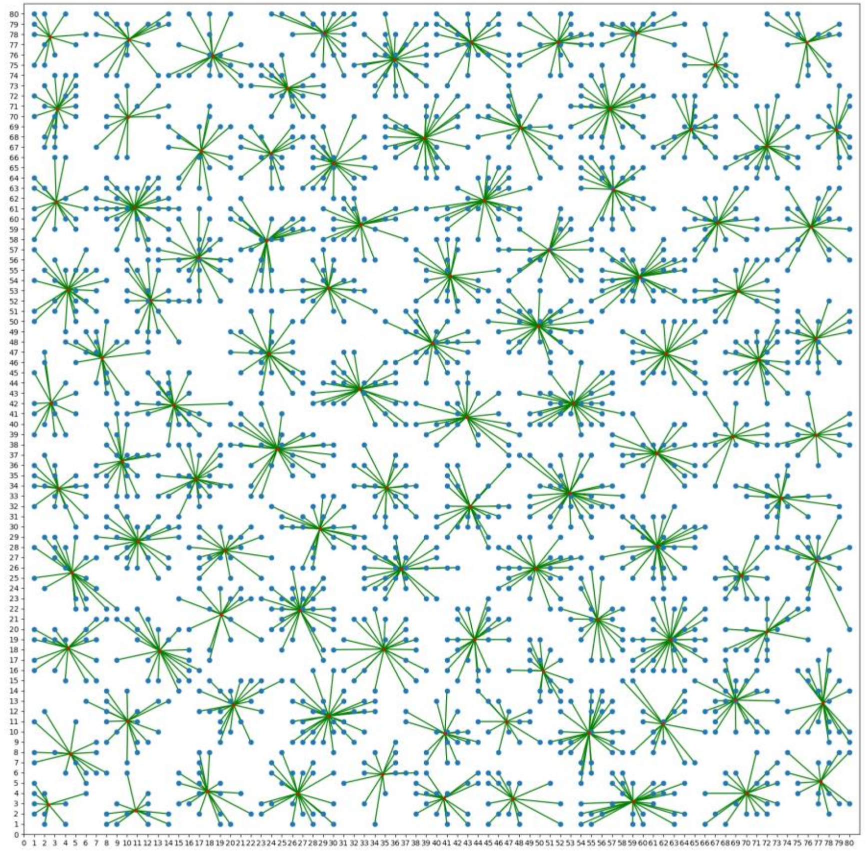
Require: S : all sites, k : The number of the center
Ensure:

```
function initial_center_selection_algorithm(S, k)
    min_s, min_value = -1, infinite
    for s in S do
        s_closed = [sites first 1/3 closest to s]
        current_max = max([dist(s, s_c) for s_c in s_closed])
        if min_value > current_max then min_s = s
        end if
    end for
    return min_s
end function
```

SUSTech Southern University of Science and Technology

INITIALIZATION ALGORITHM FOR K-MEANS





SUSTech Southern University of Science and Technology

INITIALIZATION ALGORITHM FOR K-MEANS

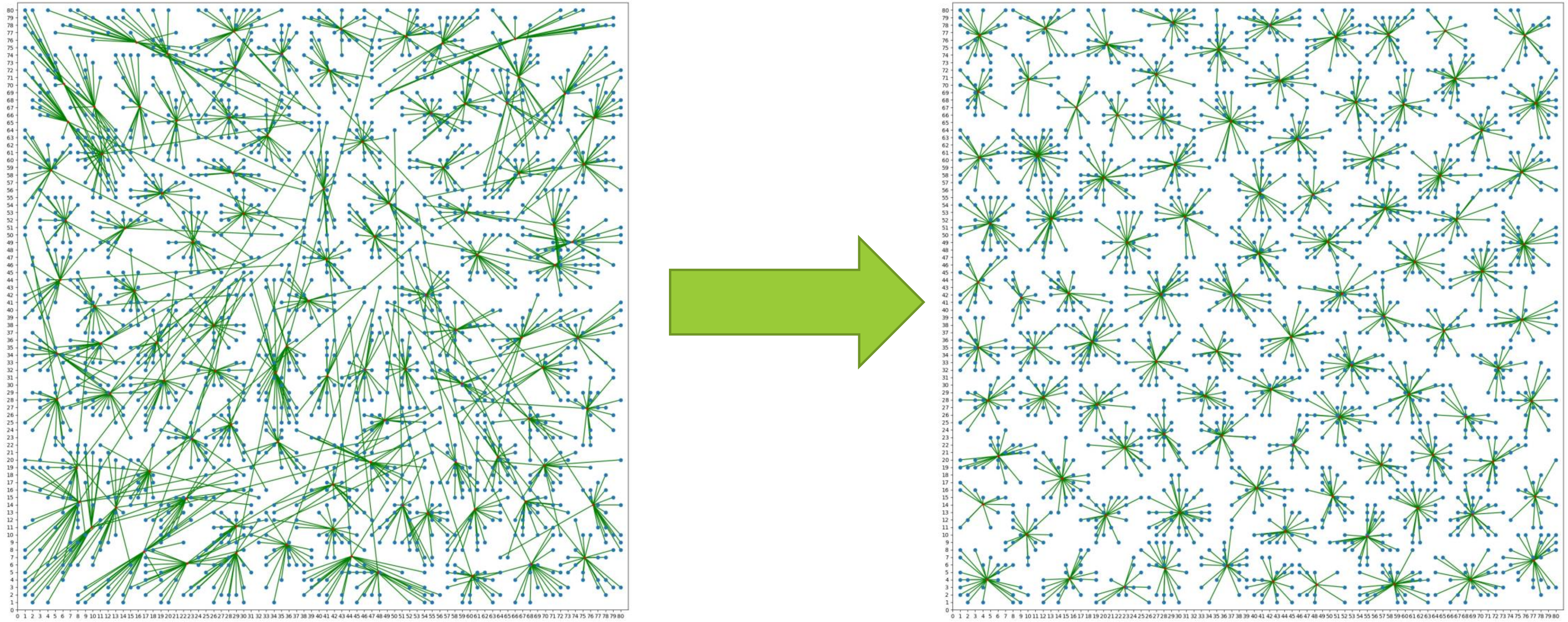
1. Initial by cluster:
   I. Use the way last page to get the k site, and let them be k partitions, each cluster has one site.
   II. Poll each partition, choice the nearest free site and add it into the partition, until all the sites have been added into the partitions.

Algorithm 7 Initial partition selection algorithm

Require: S : all sites, k : The number of the partitions
Ensure:

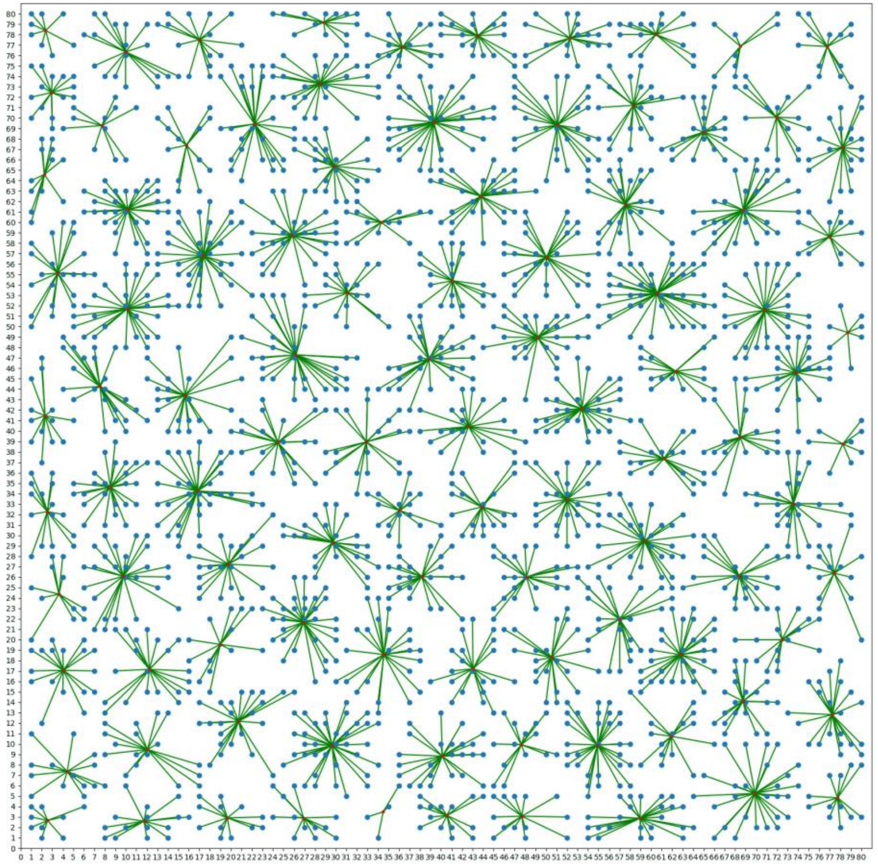
```
function initial_partition_selection_algorithm(S, k)
    initial_partition = initial_center_selection_algorithm(S, k)
    while There are sites not in partitions do
        for partition in initial_partitions do
            partition.append(nearest free site)
            if There is no free site then
                Break
            end if
        end for
    end while
    return initial_partition
end function
```

INITIALIZATION ALGORITHM FOR K-MEANS

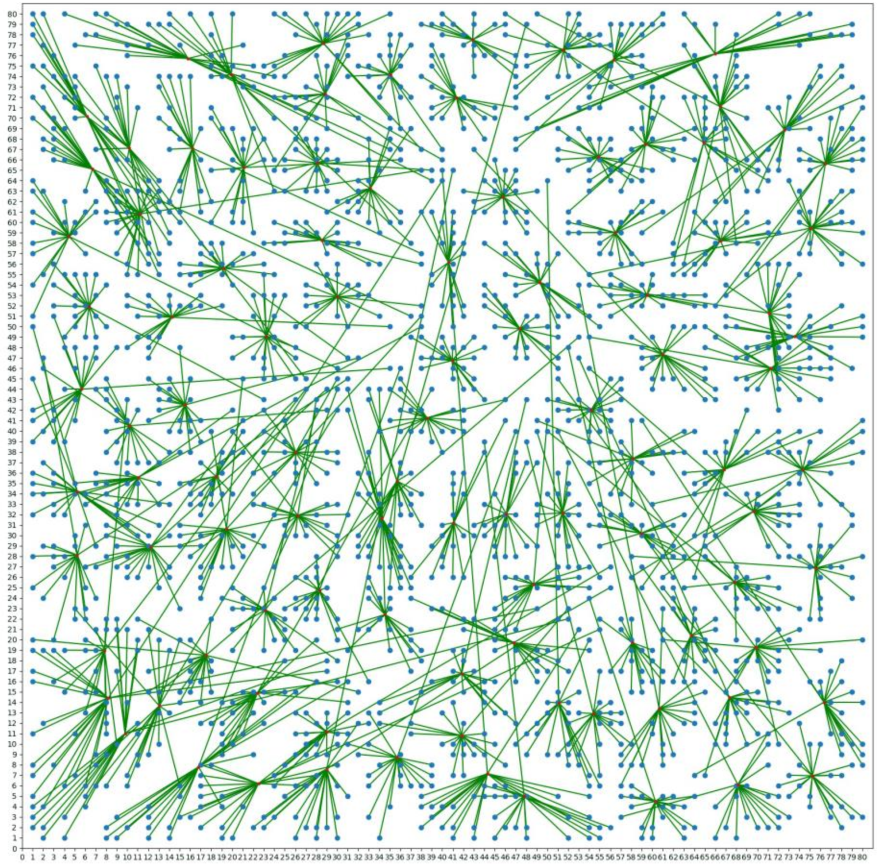


SUSTech Southern University of Science and Technology

INITIALIZATION ALGORITHM FOR K-MEANS



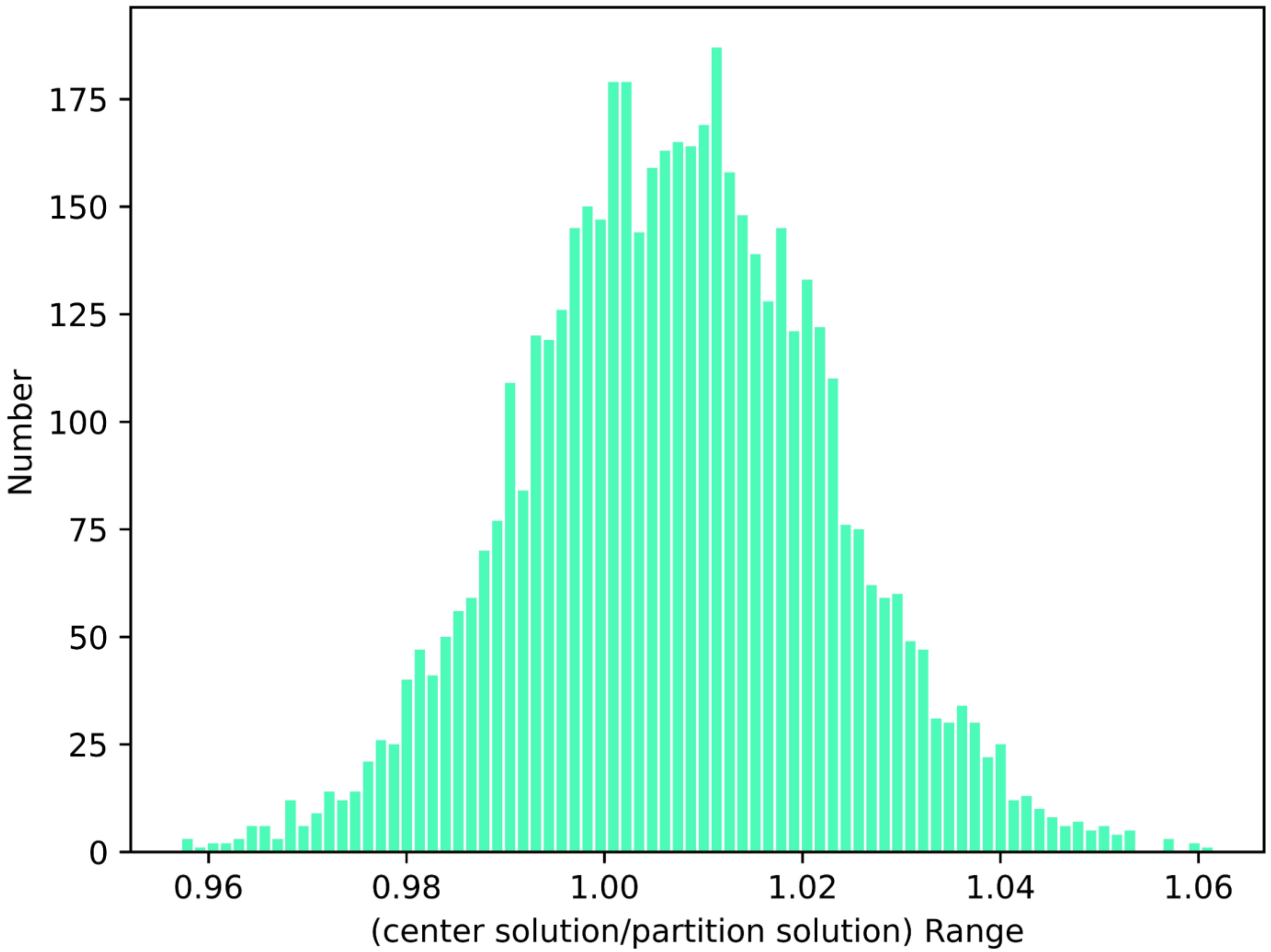
Initial by center



Initial by partition

THE ALGORITHM TO INITIALIZATION ALGORITHM FOR K-MEANS

I generate 5000 cases. Each case have 2000 sites and will be divided into 100 partitions. I do initialization for them by center and partition respectively, then I do statistic for the solution of cases.



Southern University of Science and Technology

THANK YOU !

Q&A

Southern University of Science and Technology