

Resilient k -d Trees: k -Means in Space

Soft Memory Errors

Soft memory errors

- Random bit flips, corrupting the content of the affected memory cells
- Multiple causes, e.g. power failures, alpha particles, cosmic rays
- Memories tend to be more error-prone

Occurrence rates

- Directly proportional to running time and memory used
- Seldom in individual memories, a lot in large clusters
- Increase dramatically with altitude
- **In space**: much more frequent than at ground level

k -Means

Classical k -means

- Given n input points, group them in k clusters
- **Idea**: Maintain k centers – each center is the mean of points close to it
- **Image clustering**: a cluster contains all pixels of a color, compress images
- **Motivation**: data communication between space and earth is expensive

k -means in image clustering



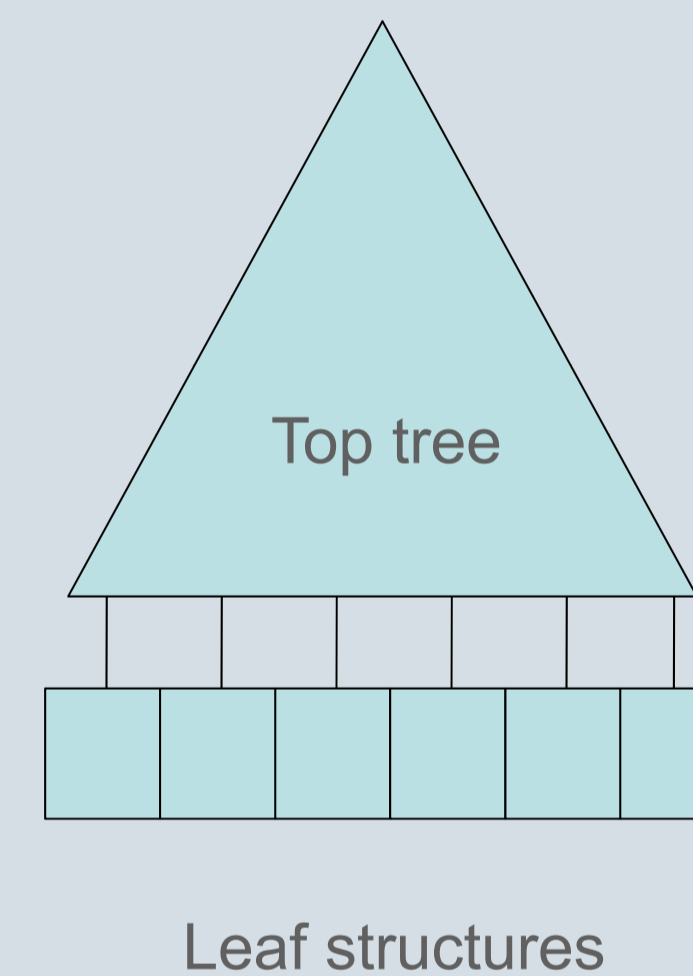
Resilient k -d k -Means

k -d k -means

- Build a k -d tree data structure on all points
- **Idea**: use the k -d tree compute the closest center for potentially large groups of points at once
- Behaves as bad as k -means on corrupted data

Resilient k -d tree

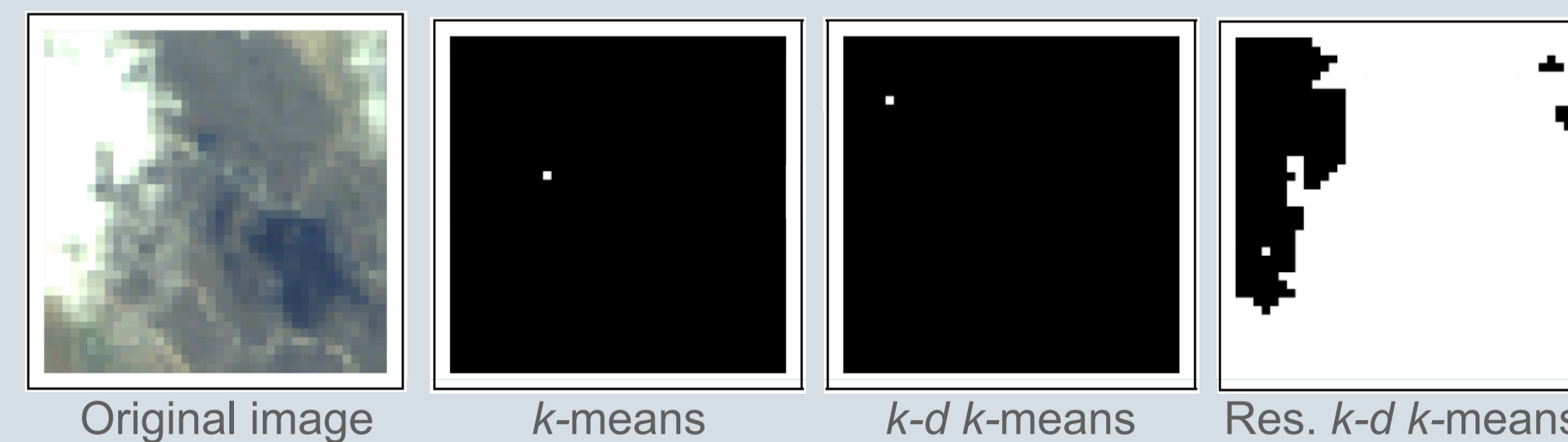
- A k -d tree guaranteed to behave well in spite of memory corruptions
- Split the tree in a **top tree** and many **leaf structures** (as in [2])
- **Top tree**: heavily replicated
- Leaf structures only contain points



Resilient k -d k -means

- k -d k -means using a resilient k -d tree
- Theoretically guaranteed to behave well on corrupted data

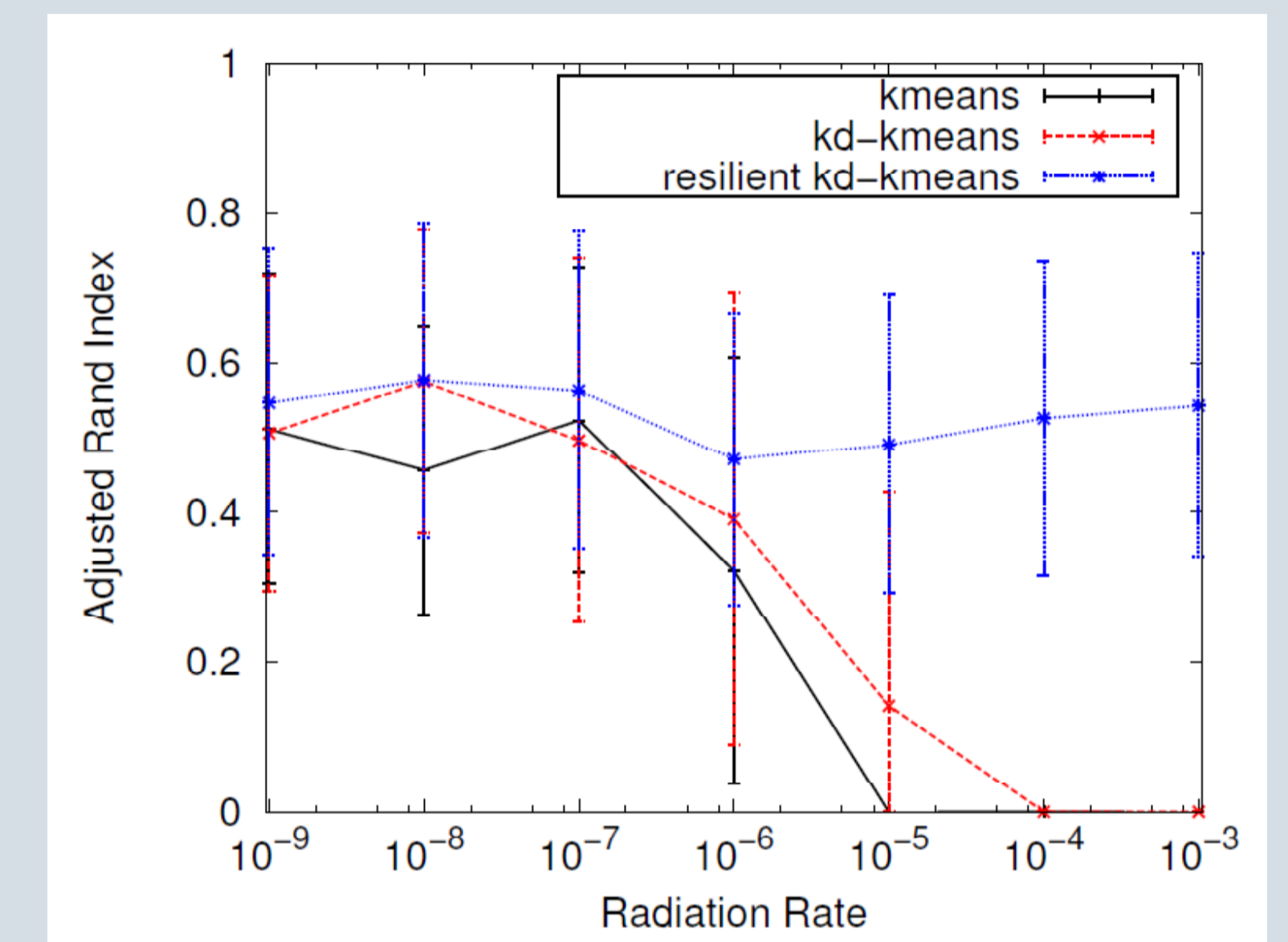
k -means in image clustering



Clustering Accuracy

Adjusted rand index

- Measure how good the clustering is
- Value of 1: perfect accuracy
- Value of 0: expected for random assignment of points to centers



Resilient k -d k -means maintains accuracy

References

[1] I. Finocchi and G. F. Italiano. *Sorting and searching in faulty memories*. In ACM STOC'04, 101–110.

[2] G.S. Brodal, R. Fagerberg, I. Finocchi, F. Grandoni, G.F. Italiano, A. Jørgensen, G. Moruz, and T. Mølhave. *Optimal resilient dictionaries*. In ESA'07, 347-358.

[3] F. Gieseke, G. Moruz, and J. Vahrenhold. *Resilient K -d Trees: K -Means in Space Revisited*. In ICDM '10.