

Conquering the Divide: Comments

- Requirement: pre-specified number of clusters
 - How to adapt the algorithms for unknown number?
- Consider noise in the data
 - Perhaps transform from deterministic clustering model to probabilistic model.
- Extended to a hierarchical setting
- Upper bound on how frequent re-clustering happens when the data changes
- Point deletion inefficiency

DisCo: Comments

- General framework
 - works for many possible co-clustering algorithms
 - works with various cost functions
- Many heuristics for speedups
 - 50-50 random split when introducing new cluster
 - Use one reducer for co-clustering part
- Sync step could be the bottleneck if G is time-consuming to compute
- Uncertain claim: "as the dataset sizes grow in future, our implementation will achieve linear scaleup."