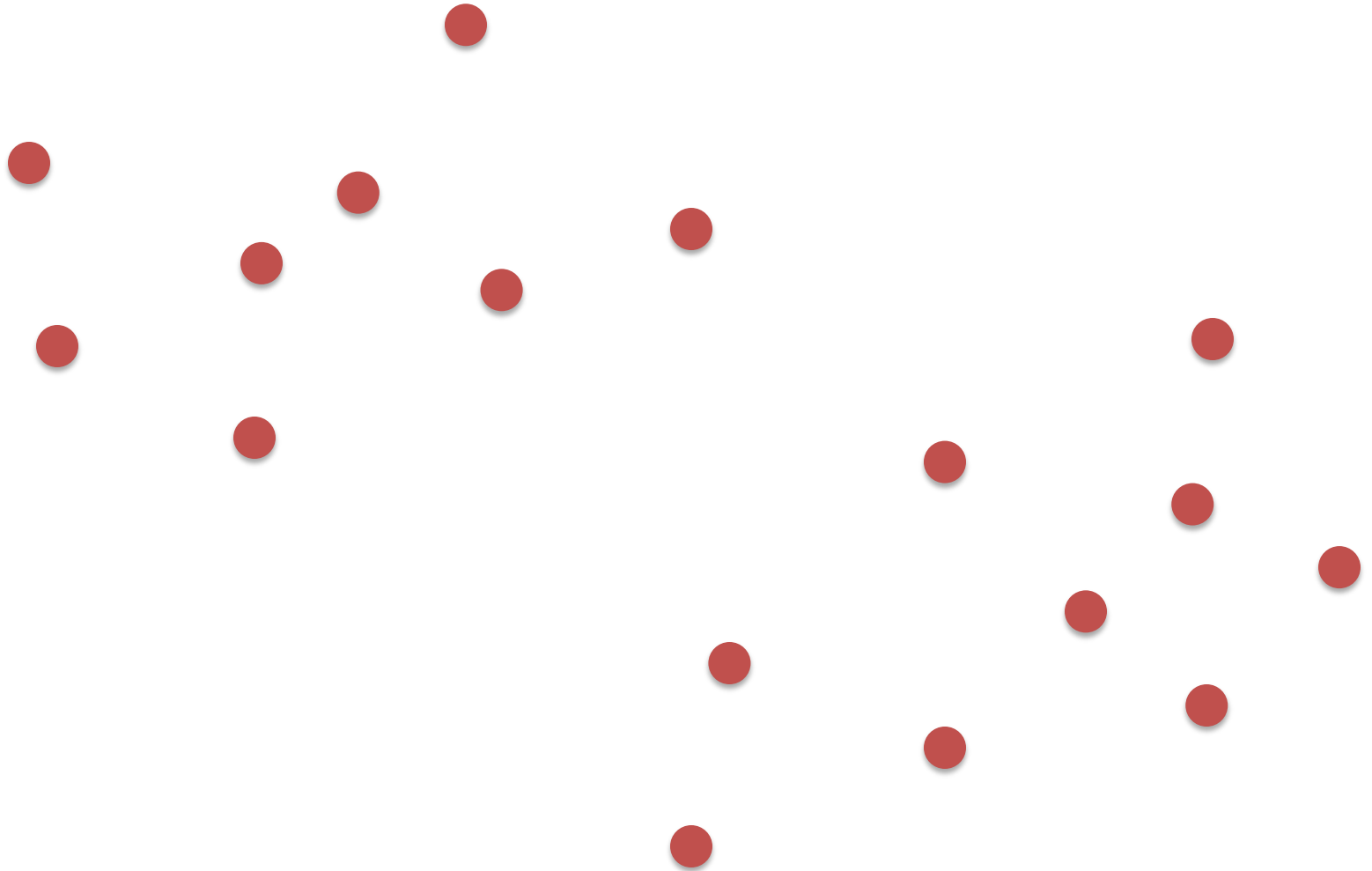
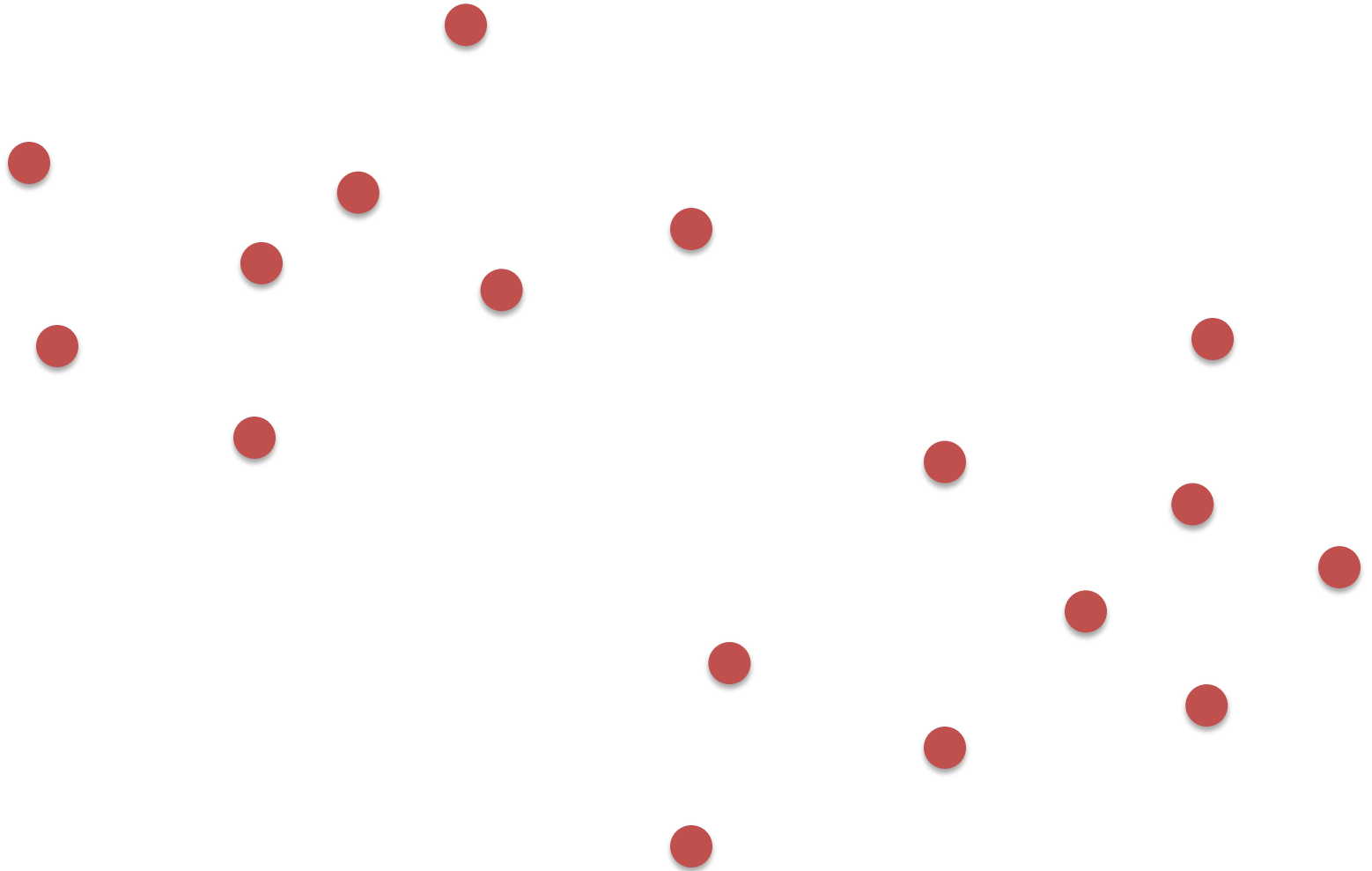


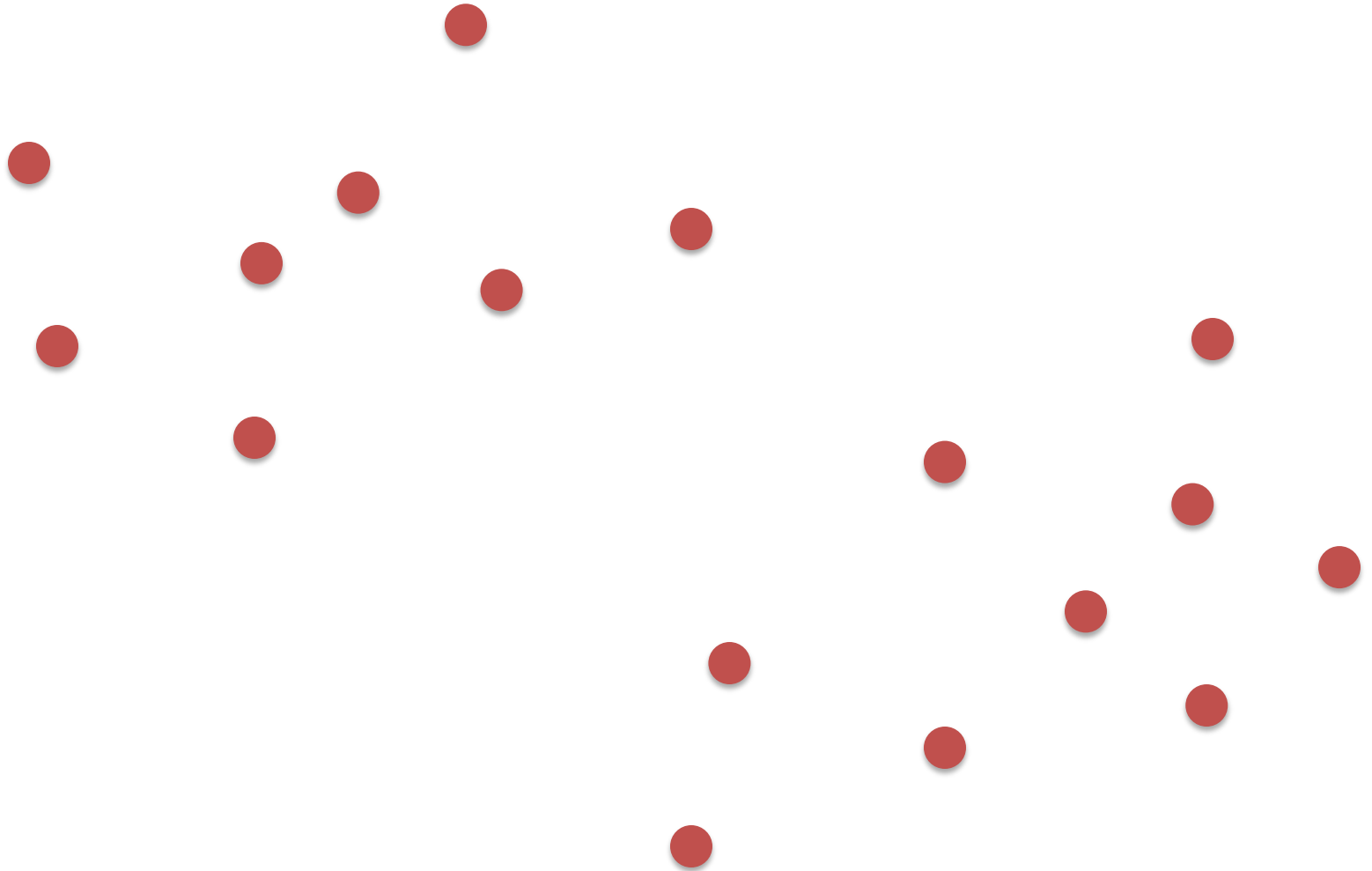
k-means



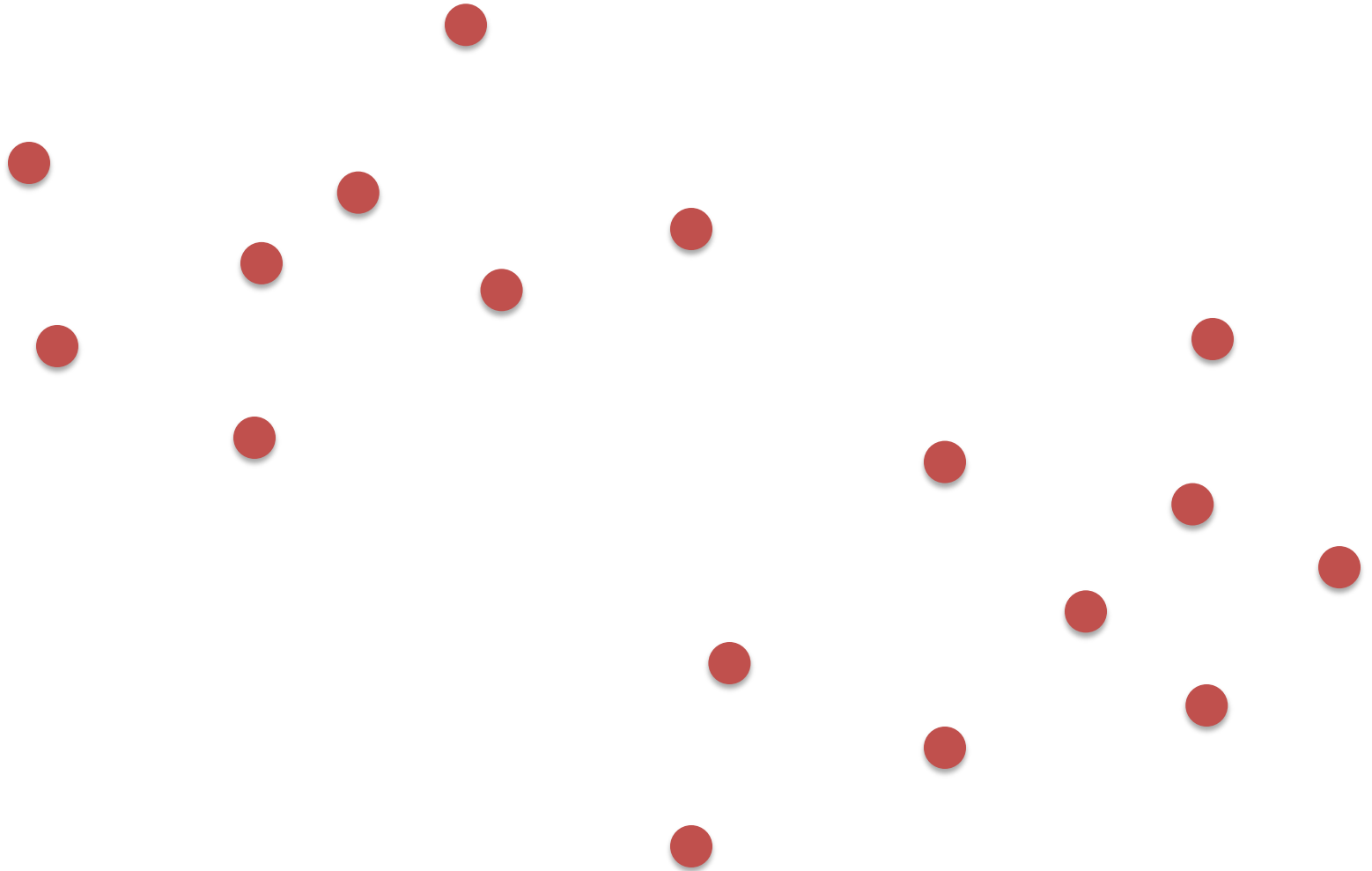
k-means



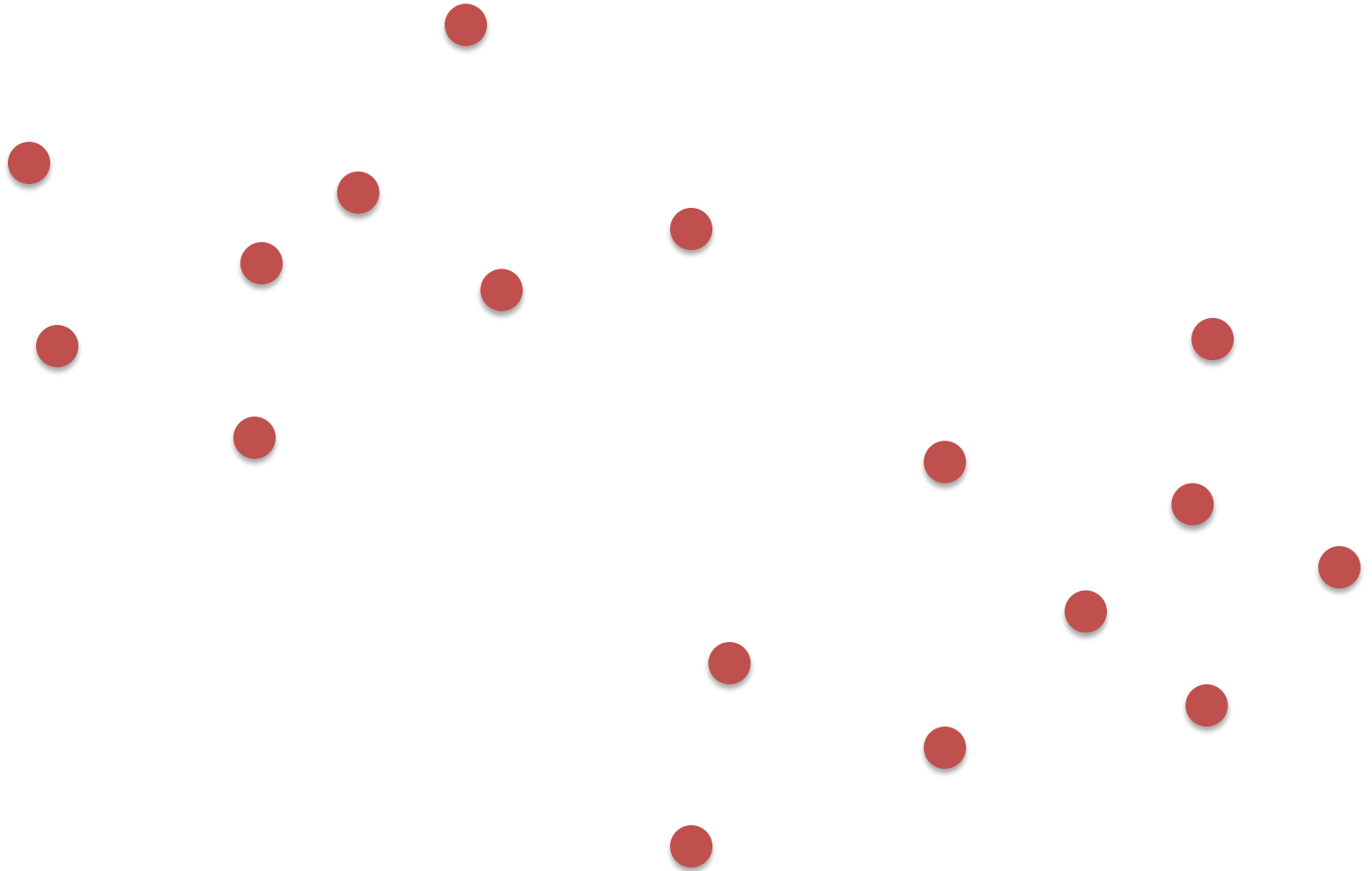
k-means



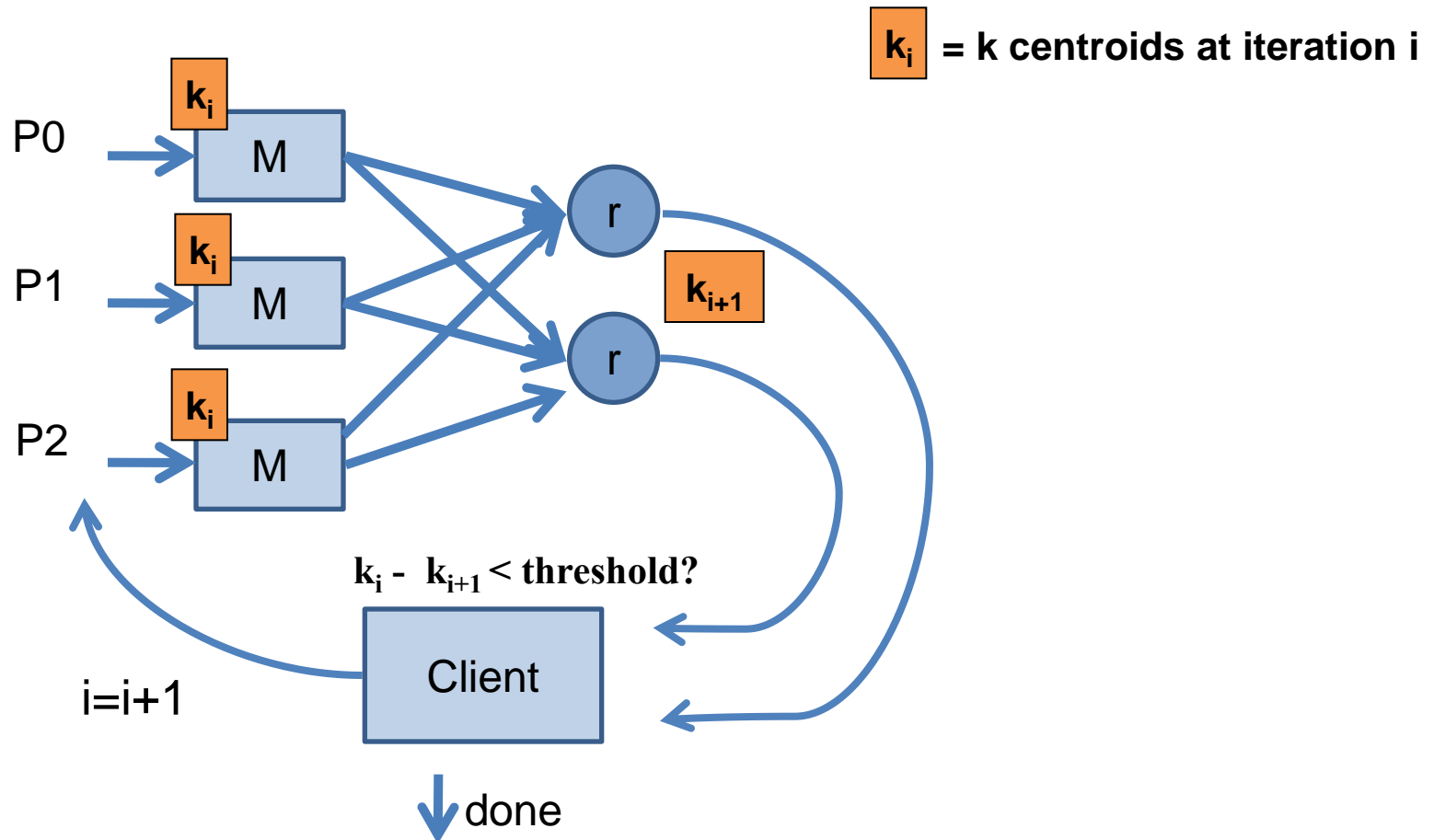
k-means



k-means



Parallelizing k-means in MapReduce



Weaknesses

- Must know the number of clusters
- Sensitivity to initial starting point
 - No unique solution
- Some sensitivity to stopping threshold