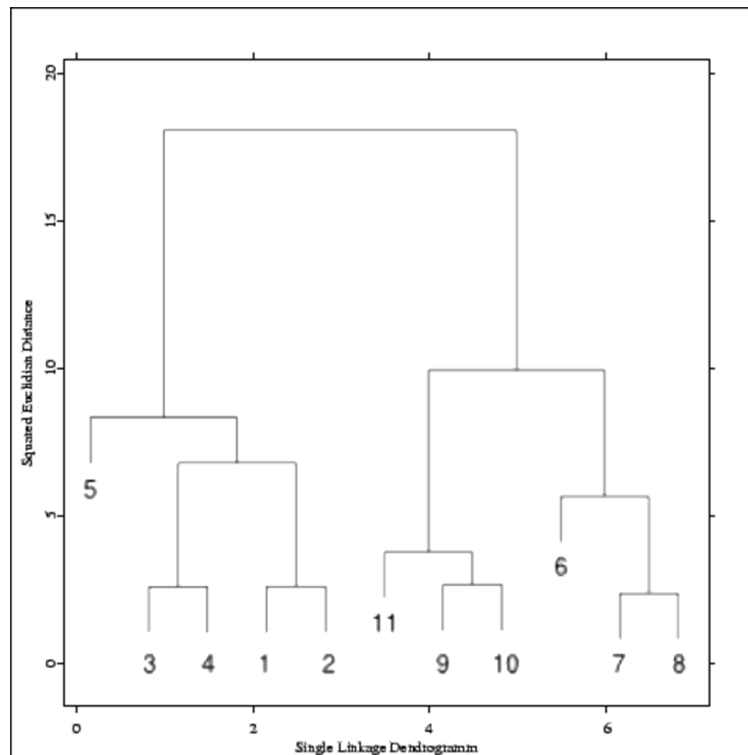


Some Review Questions:

Questions:

1. Describe clustering objectives in term of inner distance between datapoints in each cluster and distance between clusters?
2. Sum of Squared Error (SSE) is one of convergence criteria in k-mean algorithm. Explain why this measure in particular is the most common and useful comparing with the others.
3. K-mean is relatively sensitive for selecting initial centroids. Suggest an approach to overcome this shortcoming.
4. K-mean is sensitive too to outliers. Suggest an approach to overcome this shortcoming.
5. You're given the following dendrogram of a hierarchical tree clusters. Determine only 3 clusters and their associated datapoints.



6. Community finding via maximizing modularity is similar to clustering. Show some similarities and some differences.