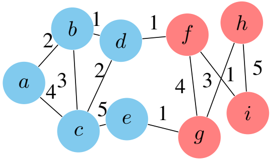
Hierarchical Clustering Algorithms

Frederik Mallmann-Trenn
6CCS3AIN

How are hierarchical clusters obtained in practice?

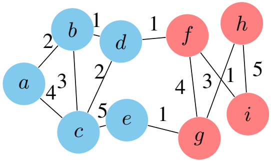
- Agglomerative clustering (bottom up)
  - Initially place each data point in its own clusters
  - Repeatedly merge most similar clusters
- Divisive clustering (top down)
  - Split using bisection k-means (or sparsest cut)
  - Recurse on each part

Agglomerative Clustering



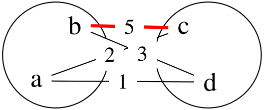
- An edge can mean similarity or dissimilarity
- In this lecture we'll only consider similarities.
- A high similarity value means that the corresponding two nodes are very similar and should be in the same cluster

Agglomerative Clustering



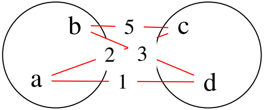
- Each node starts in a separate cluster
- The similarity between two clusters $C_1 = \{a, b\}$, $C_2 = \{c, d\}$ is

Single Linkage



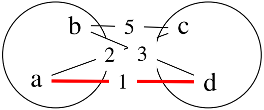
Similarity: 5

Average Linkage



Similiarity: 2.75

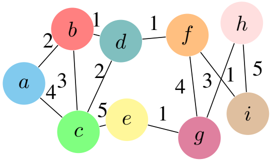
Complete Linkage



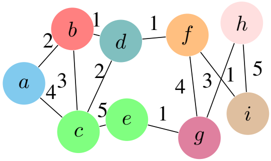
Similiarity: 1

Single-Linkage

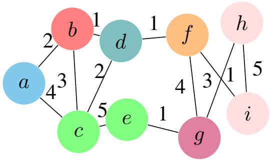
Round 0:



Round 1:

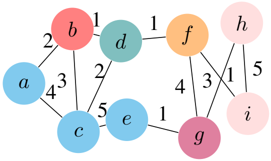


Round 2:

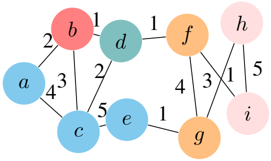


Single-Linkage

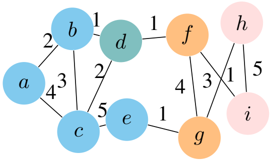
Round 3:



Round 4:

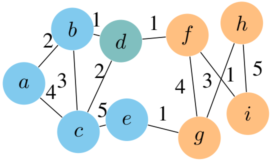


Round 5:

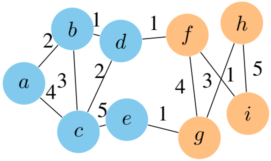


Single-Linkage

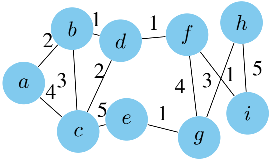
Round 6:



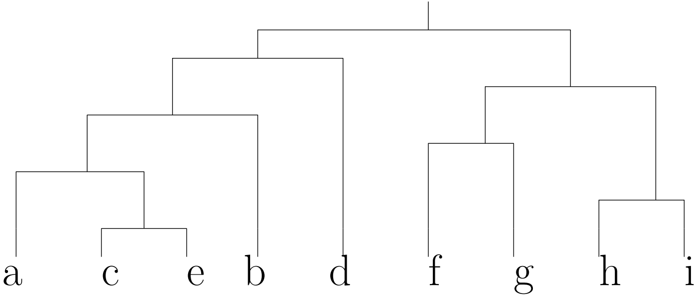
Round 7:



Round 8:

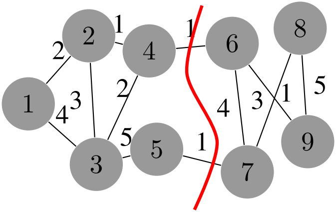


Hierarchical Clustering: Dendogram



- The obtained merges can be represented in a dendogram

Hierarchical Clustering: Divisive Heuristics



- Find a partition of the input similarity graph (or set of points)
  - Split using bisection k-means
  - Split using sparsest cut
- Recurse on each part
- Builds cluster-tree top-down

Sparsest cut

- Given a graph with nodes V and edges in $E \subset V \times V$
- The sparsity of a cut $\phi(S)$, is given by

$$\phi(S) = \frac{E(S, \bar{S})}{\min(|S|, |\bar{S}|)},$$

  where $E(S, \bar{S})$ is the sum of weights of edges crossing the cut. Here $\bar{S} = V \backslash S$.
- The sparsest cut of a graph is given by the S^* that minimises $\phi(S^*)$, i.e.,

$$\phi(S^*) = \min_S \phi(S)$$