Hierarchical Clustering Schemes

ANTHONY PM COXON
An Hierarchical Clustering scheme (HCS) (Johnson) is...

- Non-dimensional
- Hierarchical (nested inclusional subsets)
- Agglomerative [Bottom-up] (vs divisive) Clustering procedure:
- ORDINALLY INVARIANT
- Obeys Ultrametric Inequality
- UMI: \[ d_{ik} \leq \max \{ d_{ij}, d_{ik} \} \]
HCS construction: Perfect data

Figure 4.2 Illustrative example of the HCS procedure and forms of representation
HCS: noisy data

- **UMI specifies:** \( d\{[i, j], k\} = d(i, k) = d(j, k) \)

- This won’t hold for noisy data. So…

  \[ d\{[i, j], k\} = F\{d(i, k), d(j, k)\} \]

- For HCS, \( F \) is monotone function, hence HCS is non-metric:
  - \( F1 = \max \); \( F2 = \min \)

- MAX (minimizes *diameter* of cluster):
  \[ d\{[i, j], k\} = \text{MAX}\{d(i, k), d(j, k)\} \]

- MIN ("*connectedness*"):
  \[ d\{[i, j], k\} = \text{MIN}\{d(i, k), d(j, k)\} \]