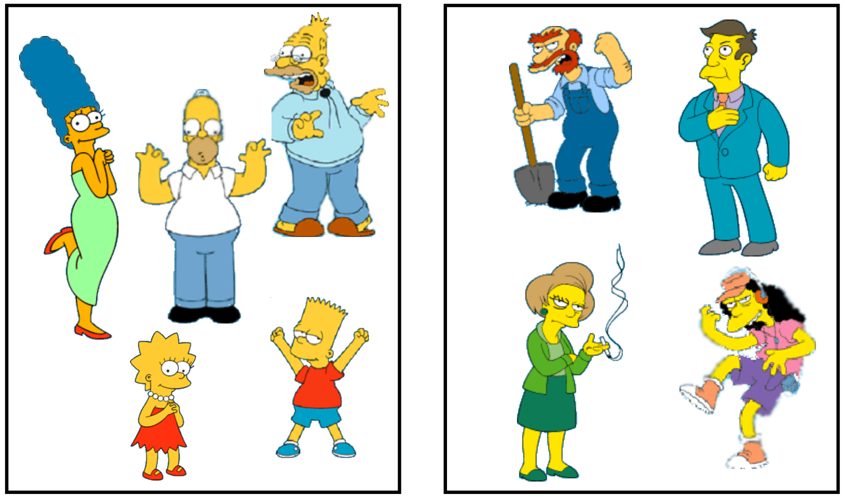
15. Clustering:
Hierarchical Clustering Algorithms

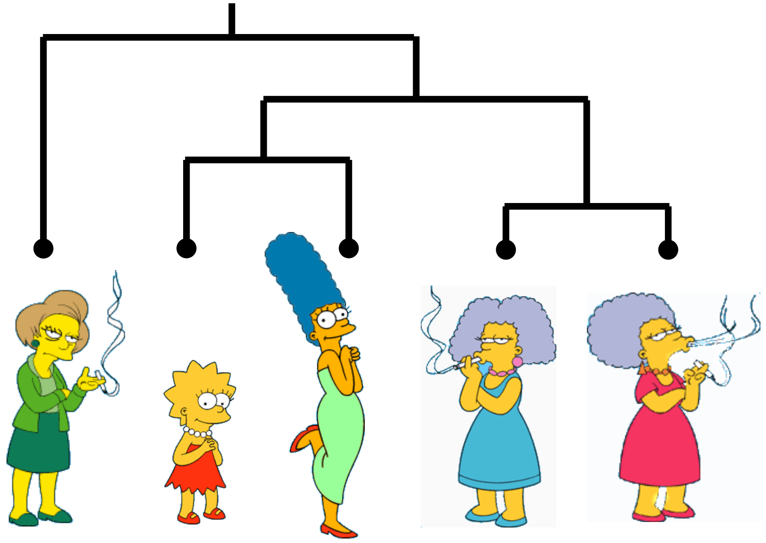
Hierarchical Clustering vs k-Means Cluste

k-Means Clustering



- It partitions the data samples onto a number of clusters
- It requires one to specify the number of clusters, distance measure

Hierarchical Clustering



- It builds a binary tree of the dataset that successively merges similar groups of data samples
- It requires one to specify a distance measure only

Hierarchical Clustering: Paradigms

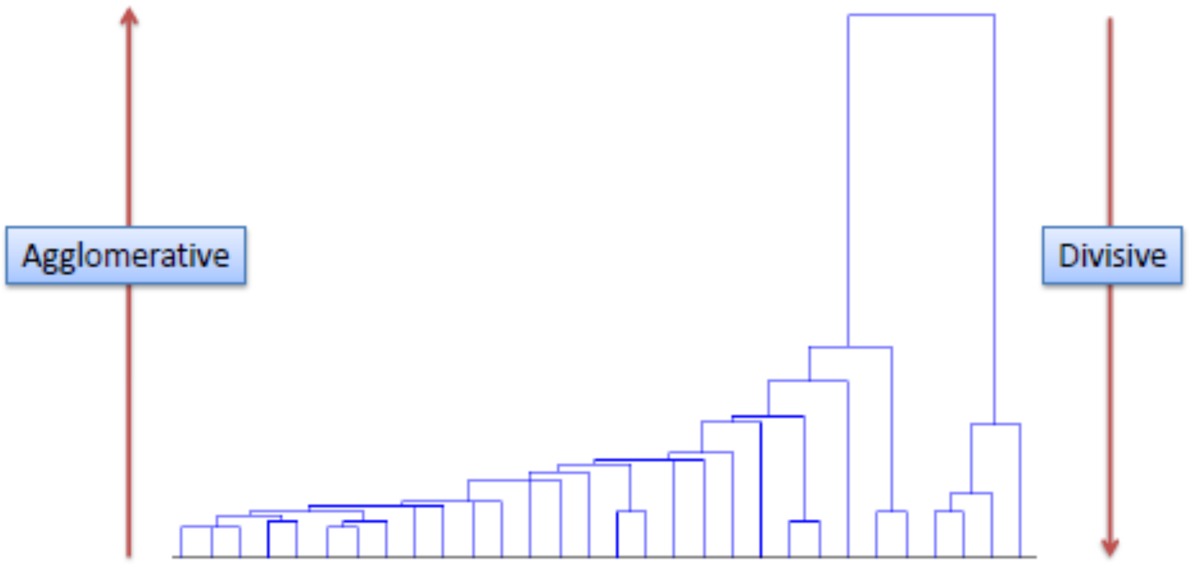
Agglomerative: Bottom-up

start at the bottom and at each level recursively merge a selected pair of clusters into a single cluster

Divisive: Top-down

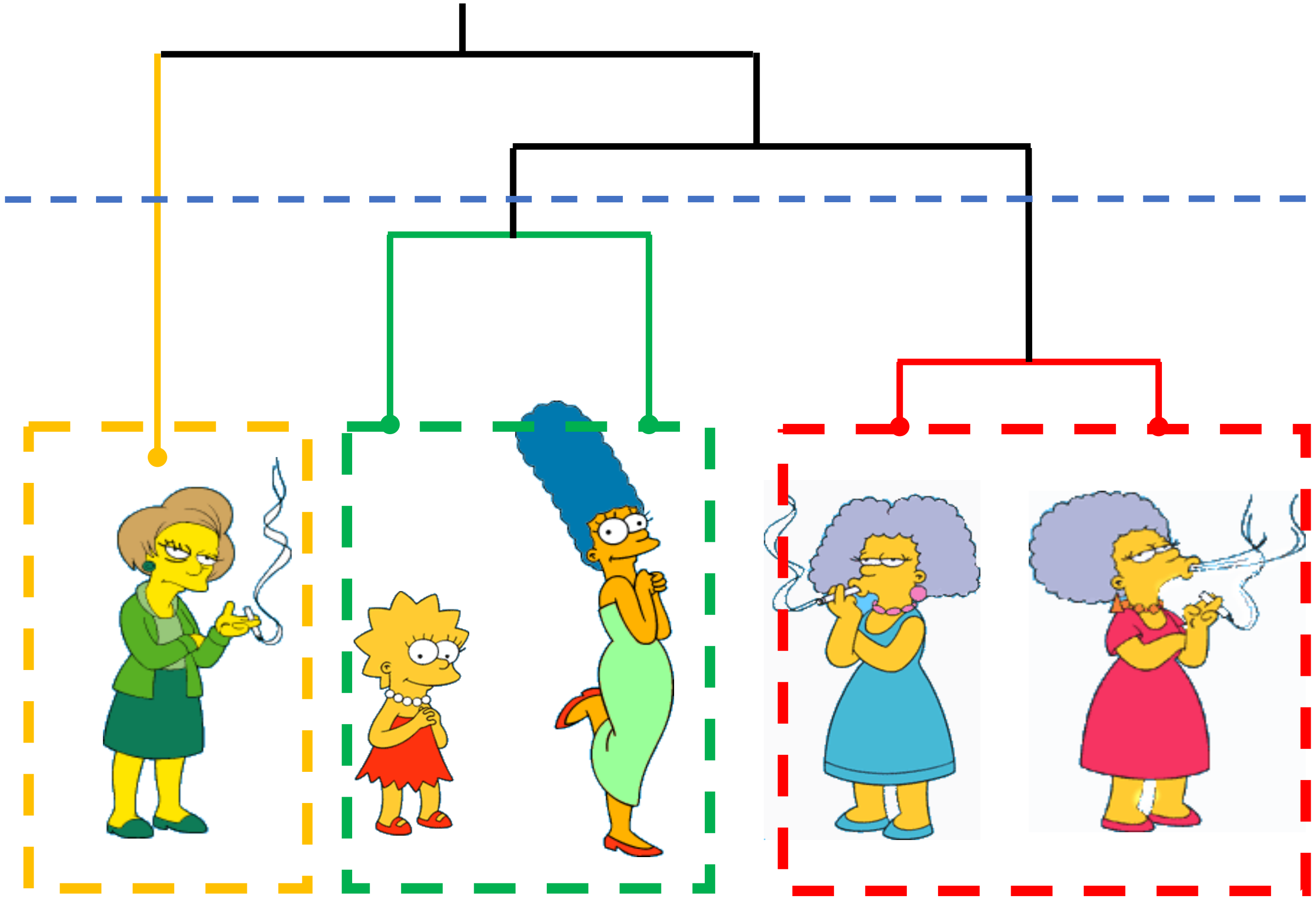
start at the top and at each level recursively split one of the existing clusters at that level into two new clusters

Hierarchical Clustering Paradigms



Dendrogram

Hierarchical Clustering: Agglomerative



Step 1: Each element is a cluster

Note:

- Simple algorithm
- Early decision cannot be undone
- Slow as at each step we merge only one new element

Hierarchical Clustering: Agglomerative

Algorithm

Input

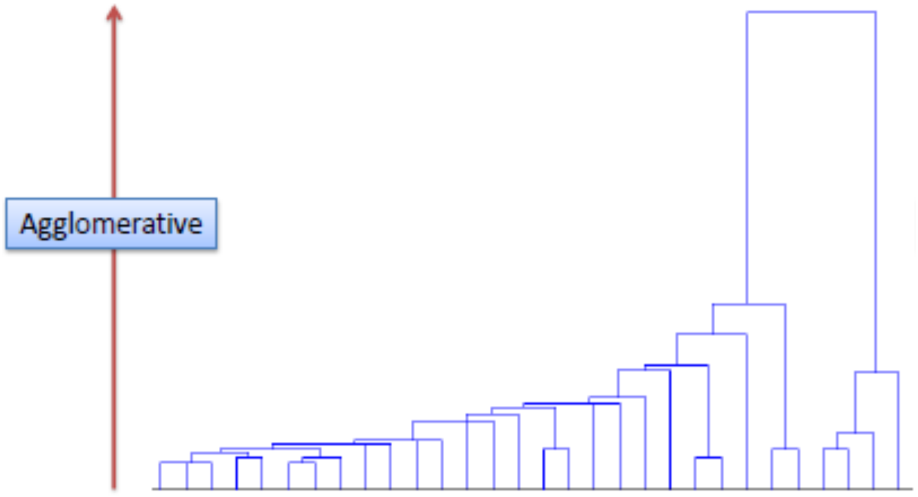
- A set of data points

Output

- A dendogram

Procedure

- Place each data point onto its own singleton group
- Repeat
  - Merge iteratively the two closest groups
- Until there is a single cluster



Similarity Measures

Single-linkage $$d(C_1, C_2) = \min_{x \in C_2, x' \in C_2} d(x, x')$$

Complete-linkage $$d(C_1, C_2) = \max_{x \in C_1, x' \in C_2} d(x, x')$$

Group average $$d(C_1, C_2) = \frac{1}{|C_1| \cdot |C_2|} \cdot \sum_{x \in C_1} \sum_{x' \in C_2} d(x, x')$$

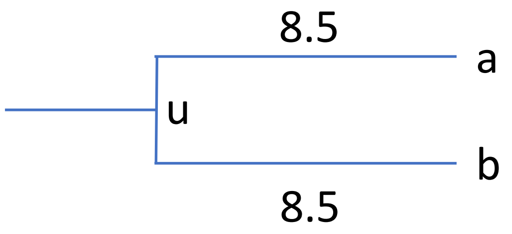
NB: The algorithm results in a sequence of groupings, but it is up to the user to choose a "natural" clustering from this sequence

Hierarchical Clustering: Example

Hierarchical Agglomerative Clustering with Group Average

Distance Matrix

	a	b	c	d	e
a	0	17	21	31	23
b	17	0	30	34	21
c	21	30	0	28	39
d	31	34	28	0	43
e	23	21	39	43	0



New Distance Matrix

	(a,b)	c	d	e
(a,b)	**0**	**25.5**	**32.5**	**22**
c	**25.5**	0	28	39
d	**32.5**	28	0	43
e	**22**	39	43	0

$$D((a,b),c) = \frac{1 \cdot D_{a,c} + 1 \cdot D_{b,c}}{1+1} = (21+30)/2 = 25.5$$

$$D((a,b),d) = \frac{1 \cdot D_{a,d} + 1 \cdot D_{b,d}}{1+1} = (31+34)/2 = 32.5$$

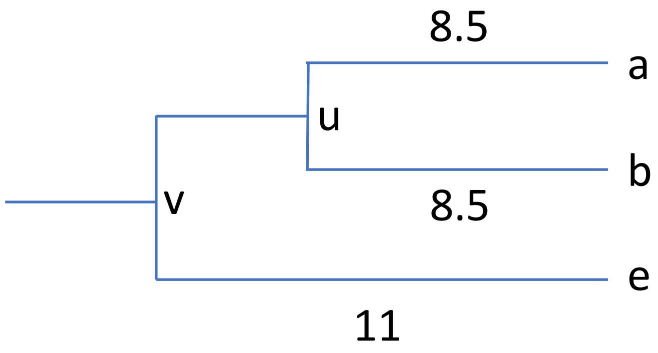
$$D((a,b),e) = \frac{1 \cdot D_{a,e} + 1 \cdot D_{b,e}}{1+1} = (23+21)/2 = 22$$

Hierarchical Clustering: Example

Hierarchical Agglomerative Clustering with Group Average

Distance Matrix

	(a,b)	c	d	e
(a,b)	**0**	**25.5**	**32.5**	**22**
c	**25.5**	0	28	39
d	**32.5**	28	0	43
e	**22**	39	43	0



New Distance Matrix

	((a,b),e)	c	d
((a,b),e)	**0**	**30**	**36**
c	**30**	0	28
d	**36**	28	0

$$D\left(\left(\left(a,b\right),e\right),c\right)=\frac{2\cdot D_{(a,b),c}+1\cdot D_{e,c}}{2+1}=\left(2\cdot 25.5+39\right)/3=30$$

$$D\left(\left(\left(a,b\right),e\right),d\right)=\frac{2\cdot D_{(a,b),d}+1\cdot D_{e,d}}{2+1}=\left(2\cdot 32.5+43\right)/3=36$$

Hierarchical Clustering: Example

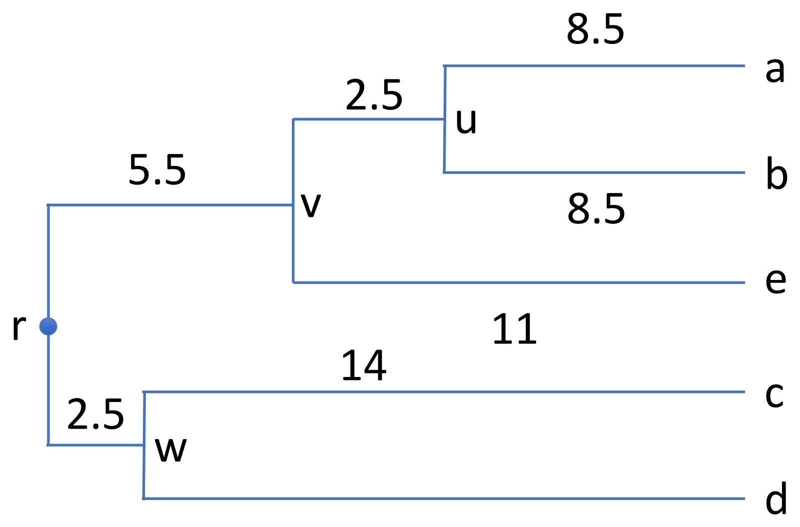
Hierarchical Agglomerative Clustering with Group Average

Distance Matrix

	((a,b),e)	c	d
((a,b),e)	**0**	**30**	**36**
c	**30**	0	28
d	**36**	28	0

New Distance Matrix

	((a,b),e)	(c,d)
((a,b),e)	**0**	**33**
(c,d)	**33**	0



$$D\big(((a,b),e),(c,d)\big)=\frac{3\cdot D_{((a,b),e),c}+3\cdot D_{((a,b),e),d}}{3+3}=(30+36)/2=33$$

Hierarchical Clustering: Example

Hierarchical Agglomerative Clustering with Group Average

Final Dendogram