HUDM5124 Session 12: Nonhierarchical clustering models

(overlapping clustering, “clumping”)

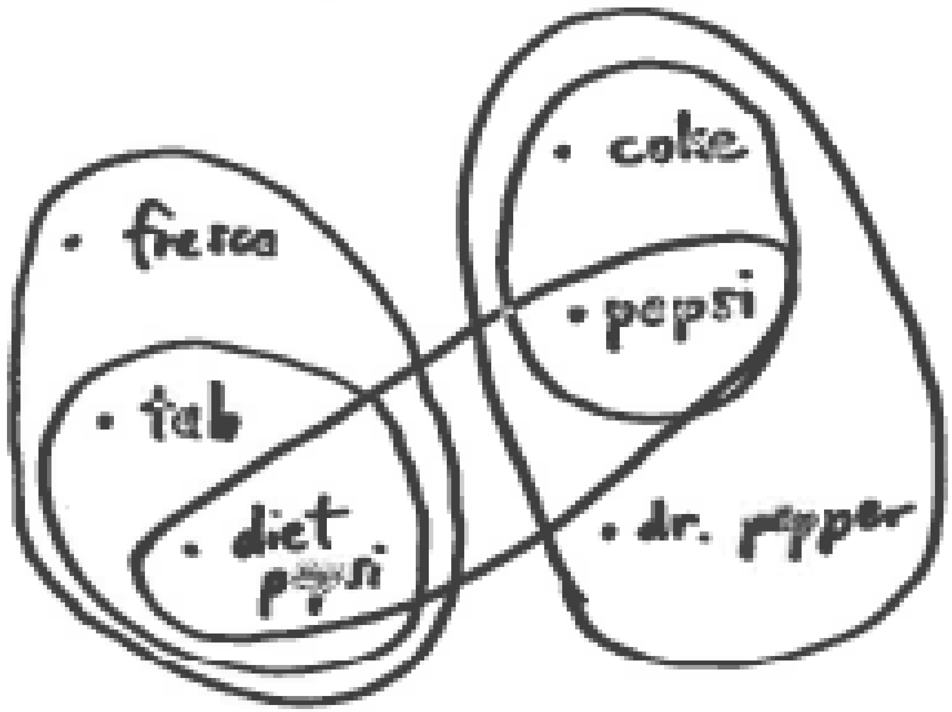
Overview:

- Additive clustering (ADCLUS)
- Extended trees (EXTREE)
- Multiple trees
- Other related models: INDCLUS, two-mode clustering, etc.
- Hybrid models (trees + spaces, etc.)

Overlapping (additive) clustering

- Additive clustering is the term most often used to denote a cluster solution in which the clusters may overlap in arbitrary patterns
- One of the first formal explorations of this idea was the ADCLUS (additive clustering) model proposed by Shepard and Arabie (1979)
- A more effective algorithm (MAPCLUS) for fitting this model was introduced by Arabie & Carroll (1980)
- Another effective approach, that can also fit the weighted INDCLUS model for individual differences, was described by Chaturvedi & Carroll (1994)

Graphically, an overlapping clustering is often represented as a set of "circles" drawn onto some spatial mapping of the object points:



Mathematically, the additive clustering model can be expressed as:

$$\hat{S}_{ij} = \sum_k w_k p_{ik} p_{jk}$$

Note that each cluster k is associated with a weight, w_k.

To express the ADCLUS model in matrix terms, note that a clustering is a collection of sets. We can represent this set of sets by an *n x m* cluster membership (or "property") matrix. The rows of this matrix correspond to conceptual "objects" to be clustered, and the columns to clusters.

P =

	C1	*C2*	*C3*	*C4*	*C5*
coke	1	1	0	0	0
diet pepsi	0	0	1	1	1
dr. pepper	0	1	0	0	0
fresca	0	0	0	1	0
pepsi	1	1	0	0	1
tab	0	0	1	1	0

Then we can write the ADCLUS model as: $\hat{S}$ = PWP'

Or as $\hat{S}$ = PWP' + C if we wish to include an additive constant

(W is a diagonal matrix of cluster weights)

Application: integers judged as abstract concepts
(from Shepard & Arabie, 1979)

Hierarchical Clustering:

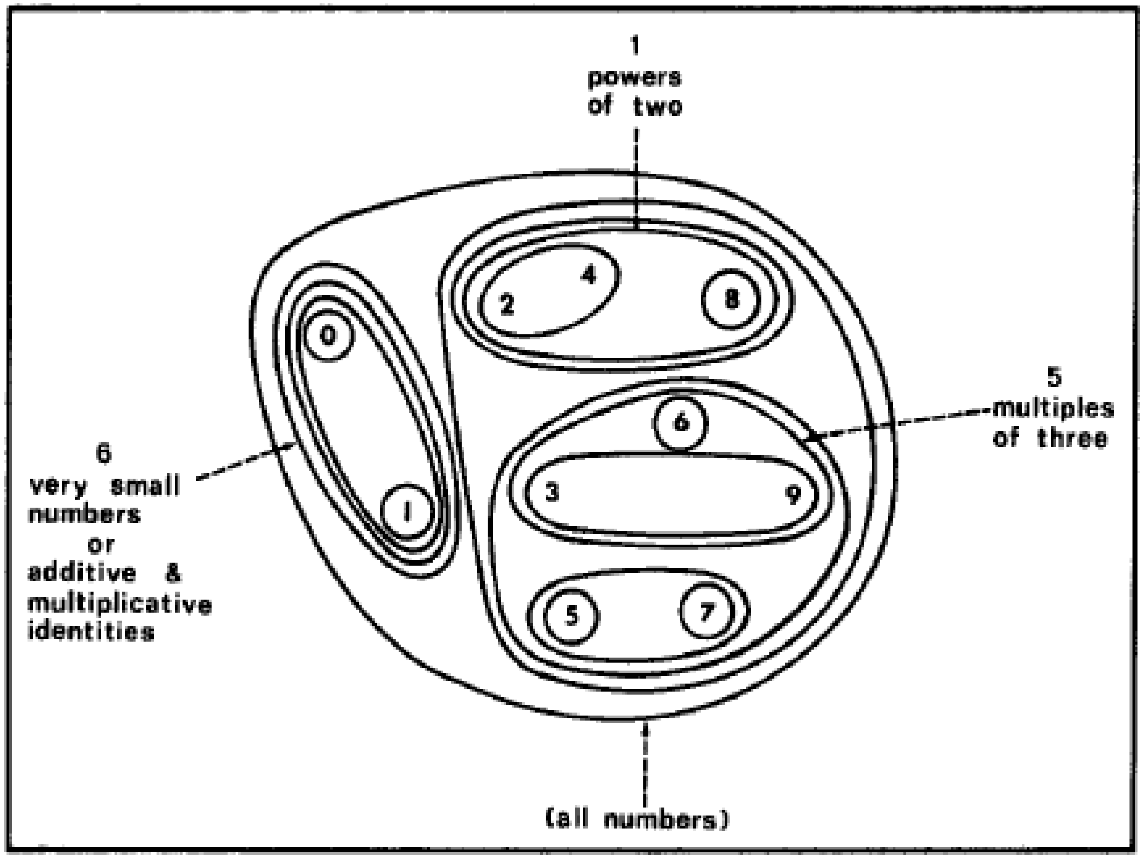


Figure 3. A strictly hierarchical clustering of the same integers embedded in the same two-dimensional scaling representation.

Application: integers judged as abstract concepts

(from Shepard & Arabie, 1979)

additive (overlapping) clustering solution:

Table 1
Judged Similarities of the Abstract Concepts of the Integers 0 Through 9

Rank[a]	*s* level	Rise	Weight	Elements of subset	Interpretation of subset
1	.638	.064	.577	2 4 8	powers of two
2	.529	.249	.326	6 7 8 9	large numbers
3	.565	.285	.305	3 4 5 6	middle numbers
4	.653	.344	.299	1 2 3	small nonzero numbers
5	.717	.344	.277	3 6 9	multiples of three
6	.574	.192	.165	0 1	additive and multiplicative identities
7	.328	.132	.150	1 3 5 7 9	odd numbers
8	.579	.206	.138	5 6 7	moderately large numbers
9	.382	.118	.112	0 1 2	small numbers
10	.235	.093	.101	0 1 2 3 4	smallish numbers

Note. The data are from Shepard, Kilpatric, and Cunningham (1975). Variance accounted for = 83.1% with 10 subsets, plus additive constant (corresponding to the complete set of 10 numbers). Additive constant = .195.

ADCLUS solution:

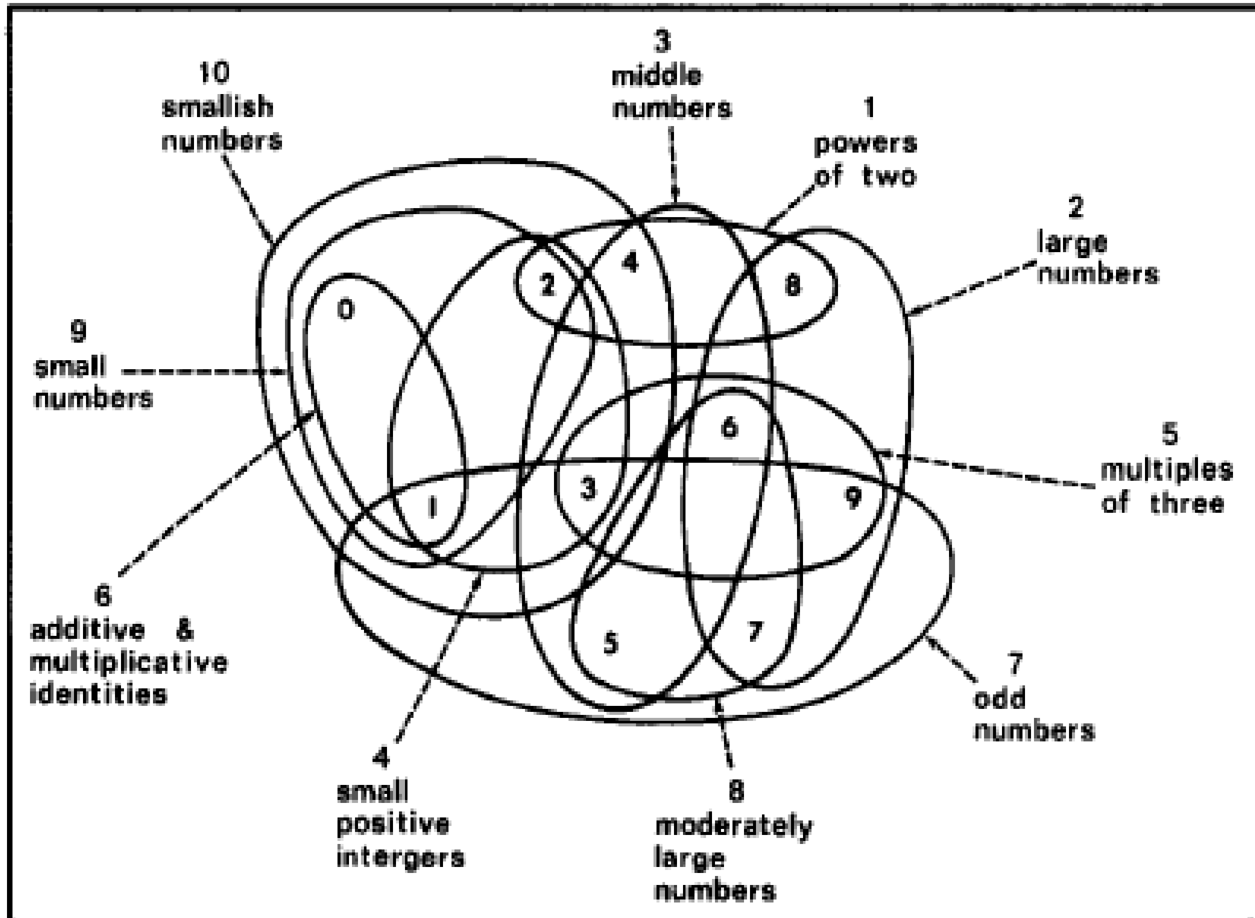


Figure 2. The 10 ADCLUS subsets obtained for the 10 integers, 0–9, studied by Shepard, Kilpatric, and Cunningham (1975), embedded in a two-dimensional scaling representation.

Software for fitting overlapping clustering

Algorithm (Model)	Authors / reference	Implementation	Available from
ADCLUS	Shepard & Arabie, 1979	FORTRAN source	??
MAPCLUS (ADCLUS)	Arabie & Carroll, 1980	FORTRAN source	NETLIB
---- (ADCLUS)	M. Lee, 2002		
SINDCLUS (ADCLUS, INDCLUS)	Chaturvedi & Carroll (1994)	FORTRAN source	NETLIB
SYMPRES (INDCLUS)	Kiers, 1997 (see also Wilderjans et al., 2012)	??	??
	Lee, M. D. (1999)		

Newer approaches to fitting the ADCLUS model Michael D. Lee

Lee, M. D. (1999). An extraction and regularization approach to additive clustering. *Journal of Classification, 16*, 255-281.

Lee, M. D. (2002). A simple method for generating additive clustering models with limited complexity. *Machine Learning, 49*, 39–58.

The additive clustering model can be thought of as a special case of the contrast model (Tversky, 1977), i.e. as an additive common-features model.

Ultrametric trees are also a special case of common-features models (in fact, they are additive clusterings with the additional restriction that clusters must be nested)

Additive trees can be seen as a special (nested) case of an additive distinctive-features model.

Classification of Additive Feature Models (Corter & Tversky, 1986)

		Feature Structure	
		Nested	Non-nested
Distance Rule	Common Features	Hierarchical Clustering (Sokal & Sneath, 1963)	Additive Clustering (Shepard & Arabie, 1979)
	Distinctive Features	Additive Tree (Sattath & Tversky, 1977)	Extended Tree (Corter & Tversky, 1986)

Extended Similarity Trees
(=general case of a distinctive-features model)

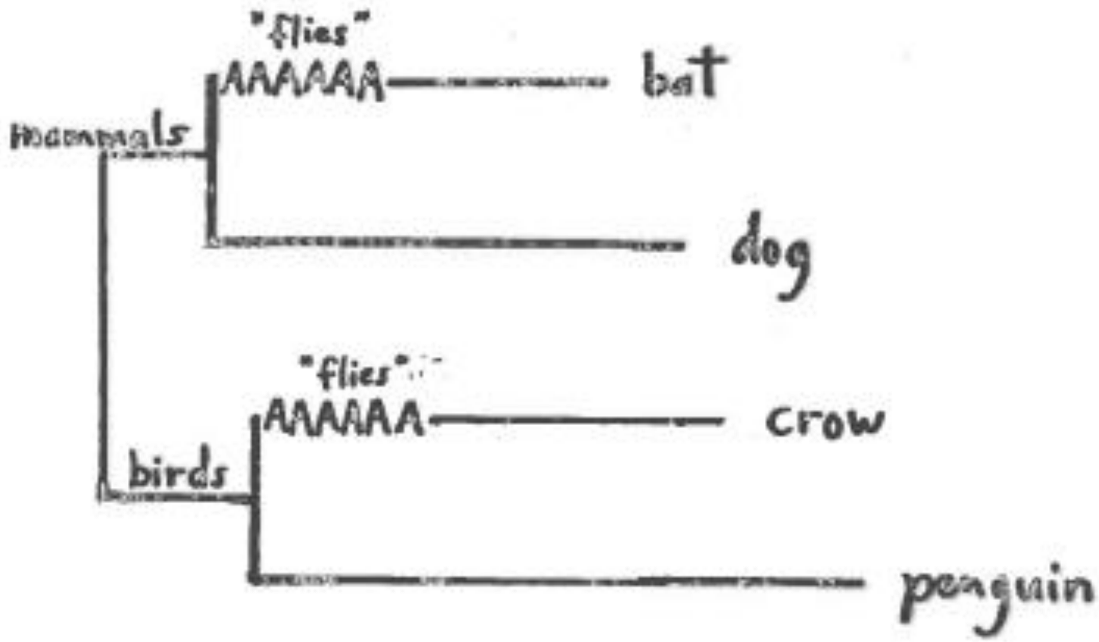
$$d(x,y) = -\theta \underbrace{f(X \cap Y)}_{\text{"common features" of } x \,\&\, y} + \alpha \underbrace{g(X-Y)}_{\text{"distinctive features" of } x} + \beta \underbrace{g(Y-X)}_{\text{"distinctive features" of } y}$$

$\longrightarrow$ interval scale of (dis)similarity

<u>Symmetric Difference Metric</u>:

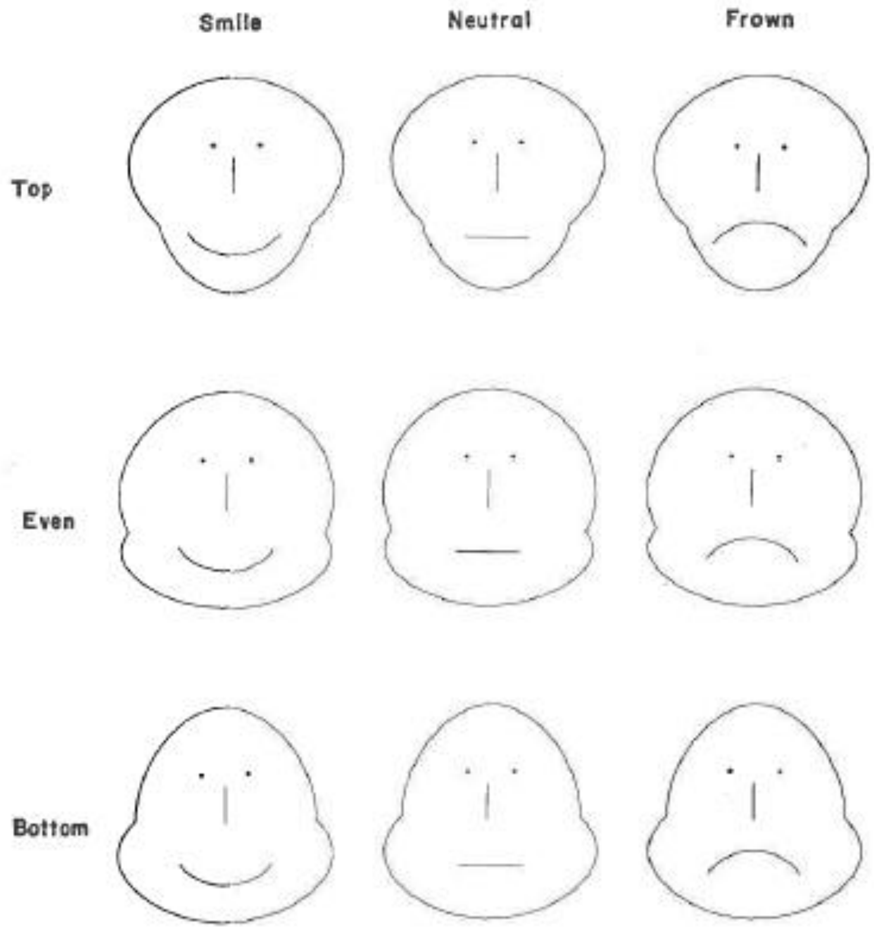
$$d(x,y) = g(X-Y) + g(Y-X)$$

An example of non-nested feature sets

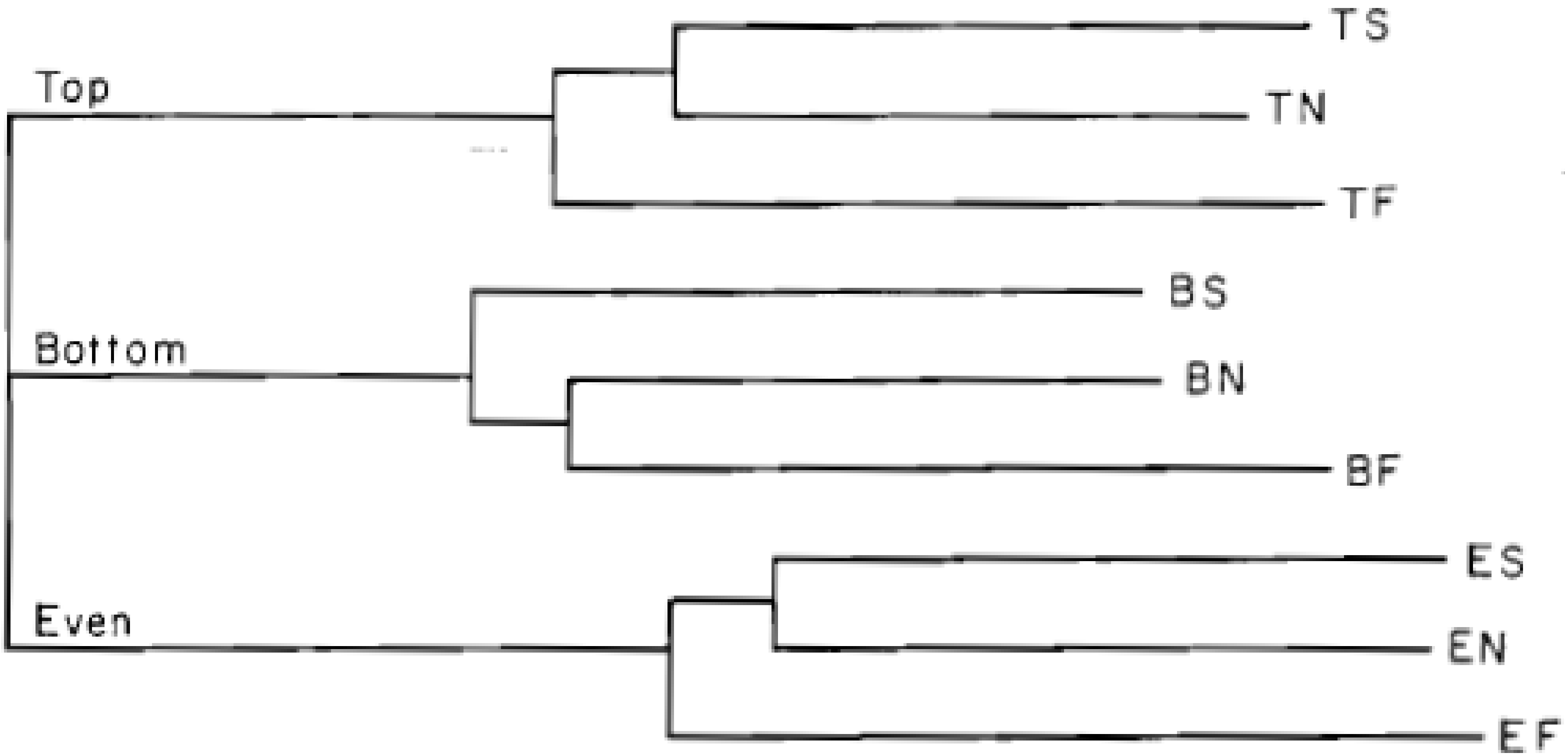


$$d(\text{bat}, \text{dog}) + d(\text{crow}, \text{penguin}) < d(\text{bat}, \text{crow}) + d(\text{dog}, \text{penguin}) < d(\text{bat}, \text{penguin}) + d(\text{dog}, \text{crow})$$

Example from Corter & Tversky (1986) – a factorial stimulus structure



ADDTREE/P solution for schematic faces (similarity ratings)



R-SQ = .58

EXTREE solution for schematic faces (similarity ratings)

R-SQ = .99

Extended Similarity Trees (EXTREE) - Corter & A. Tversky (1986)

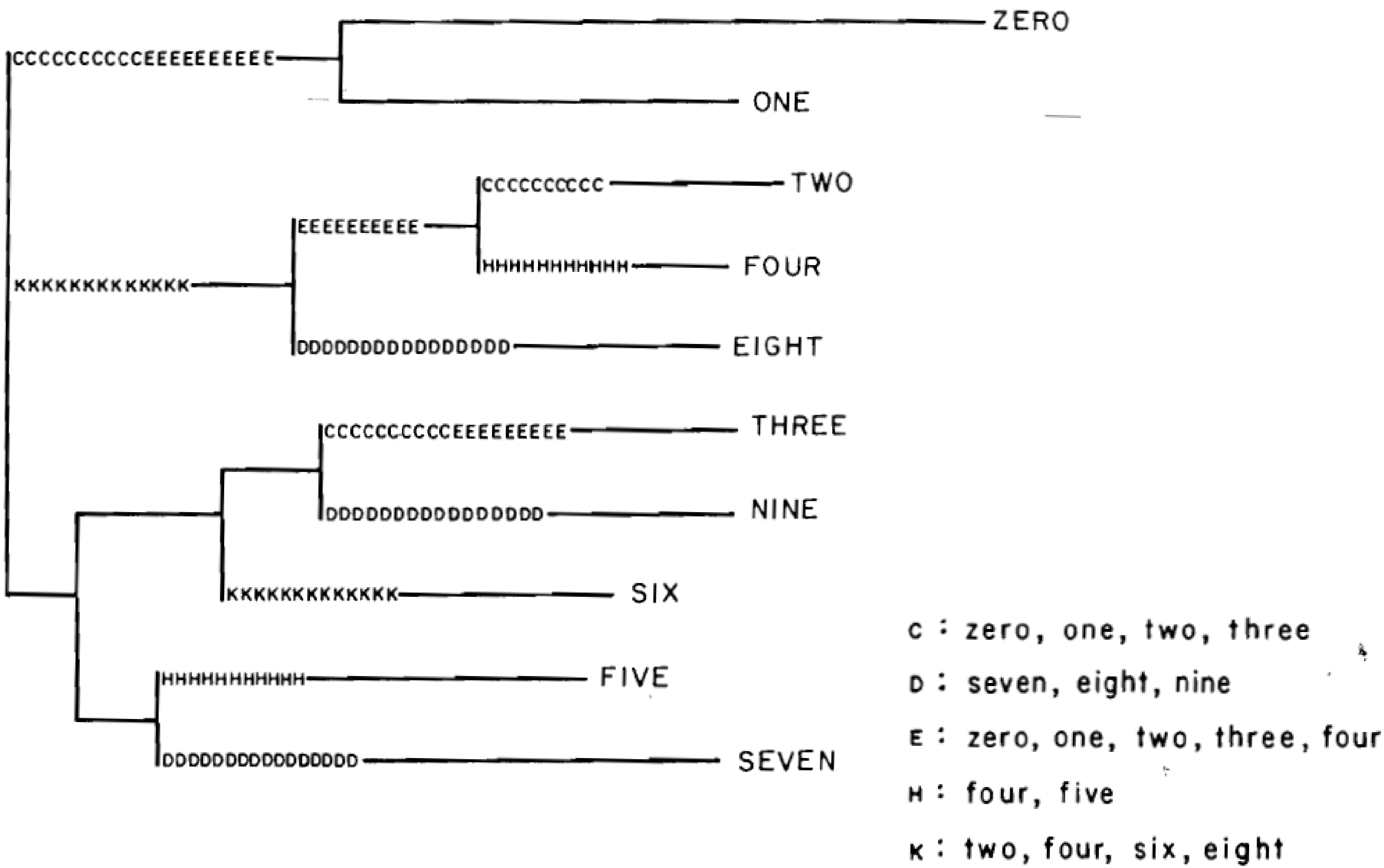


FIGURE 8
EXTREE solution for similarity of integers (Shepard et al., 1975). PV = 90%.

(high-level) Summary of EXTREE algorithm

I. fit best additive tree

- transform data to satisfy metric axioms
- D → compute neighbor score matrix, N
- combine those objects that are mutual nearest-neighb..

II. estimate measure of each possible marked feature & choose best set

$$w_A = \frac{1}{2N} \sum_{Q} \left[d(x,v) + d(y,u) - d(x,u) - d(y,v) \right]$$

- eliminate redundant features
- find cliques of pairwise features

III. simultaneous least-squares estimation of all model parameters

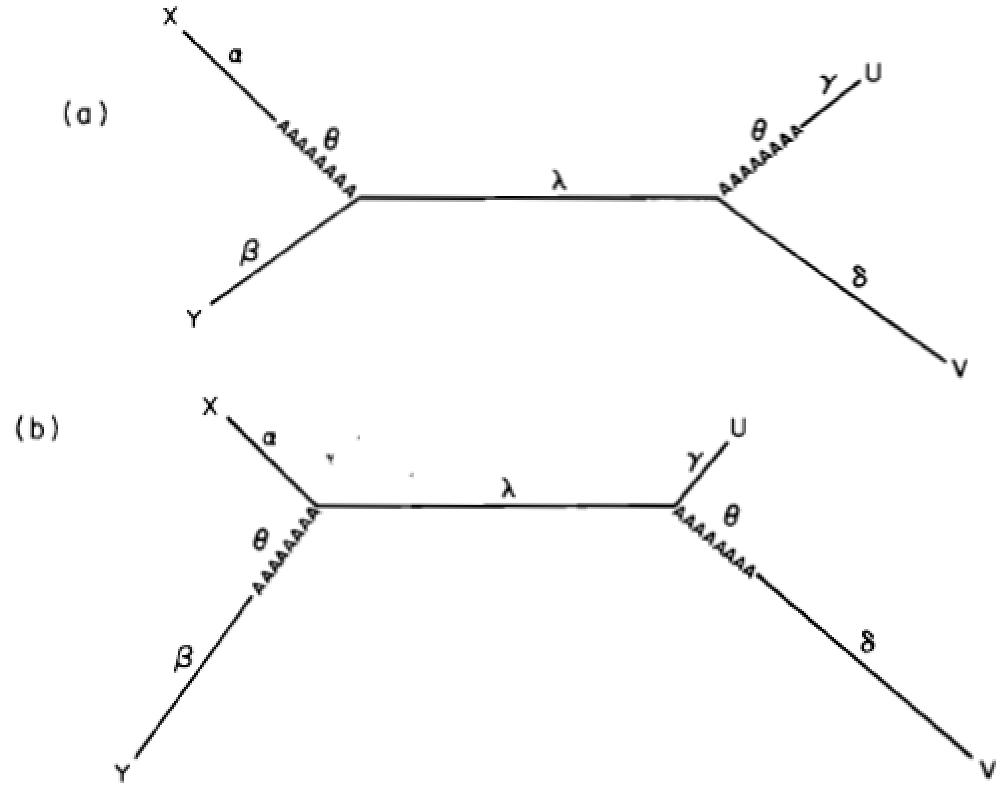
Some indeterminacies in the representation

1) a feature might be represented by tree structure OR by a marked segment

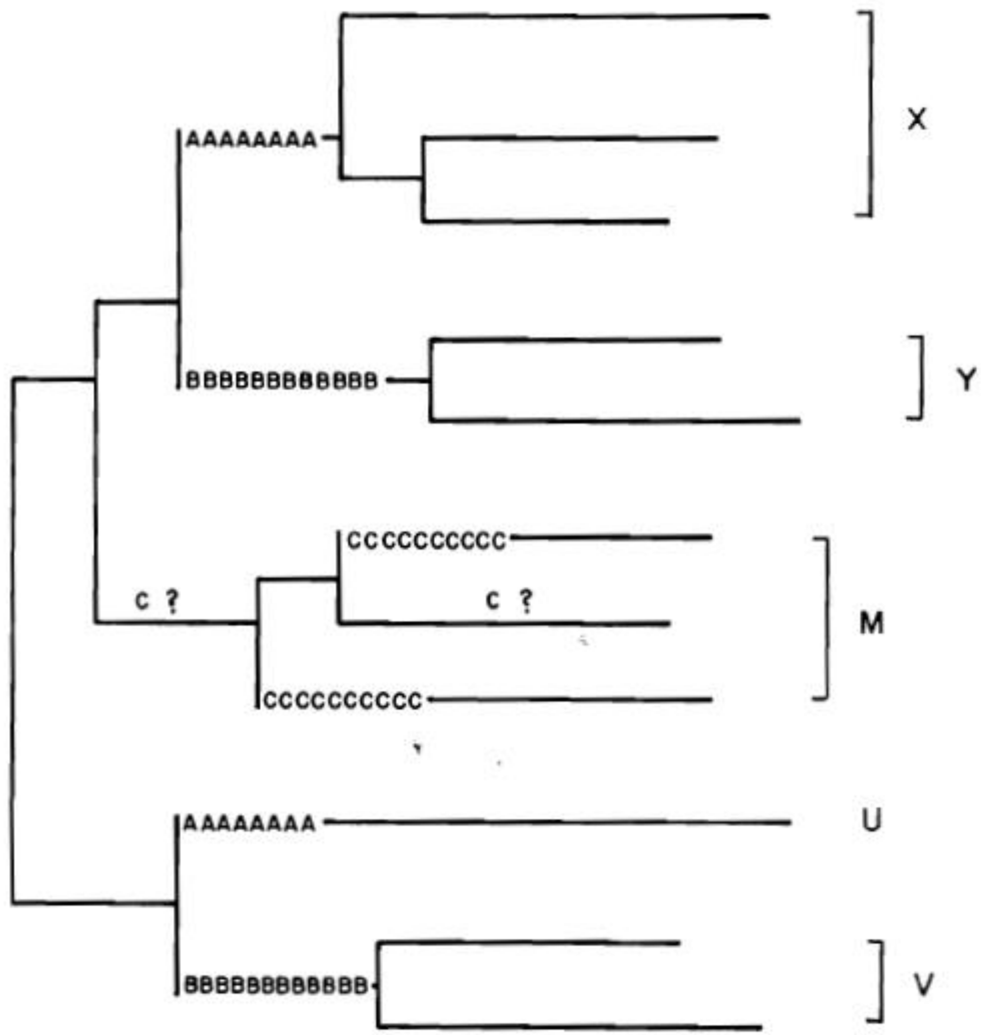
2) Placement of a marked feature (see figure)

Note that both representations are equivalent in terms of the model distances:

$d_{xy} + d_{uv} < d_{xu} + d_{yv} < d_{xv} + d_{yu}$



Some placement indeterminacies are resolved by choice of a root:



EXTREE example: numbers rated as abstract concepts (Shepard, Kilpatric, & Cunningham; 1975)

```
                                                     marked feature
                                                     pattern matrix
                                                     --------------

                --------------------------------  zero   C D . . .
CCCCCCDDDDDDD---|                                        . . . . .
|               --------------------  one                C D . . .
|                                                        . . . . .
|                      CCCCCC---------  two              C D . . N
|             DDDDDDD--|                                 . . . . .
|NNNNNNNN-----|        IIIIIII-----  four                . D . I N
|             |                                          . . . . .
|             EEEEEEEEE------------  eight               . . E . N
|                                                        . . . . .
|              CCCCCCDDDDDDD--------  three              C D . . .
|          ----|                                         . . . . .
|  --------|   EEEEEEEEE------------  nine               . . E . .
|  |       |                                             . . . . .
---|       NNNNNNNNN---------  six                       . . . . N
   |                                                     . . . . .
   |   IIIIIII---------------  five                      . . . I .
   ----|                                                 . . . . .
       EEEEEEEEE-------------------  seven               . . E . .
```

r-squared (p.v.a.f.)=0.9013

Representing overlapping clustering by multiple trees

Carroll, J. D., & Corter, J. E. (1995). A graph-theoretic method for organizing overlapping clusters into trees, multiple trees, or extended trees. *Journal of Classification*, *12*(2), 283-313.

Basic ideas:

- An overlapping clustering (i.e., ADCLUS model solution) can be grouped into two or more nested subset of features
- This can be accomplished by a clique-finding algorithm operating on a graph representation of the feature set relationships
- Each nested subset of clusters then can be represented by a tree
- The dissimilarity between any two objects is then modeled visually by the SUM of the distances in the two trees (common feature interpretation also possible)

Representing overlapping clustering by multiple trees (cont.)

Example: MAPCLUS solution for the Rosenberg kinship data

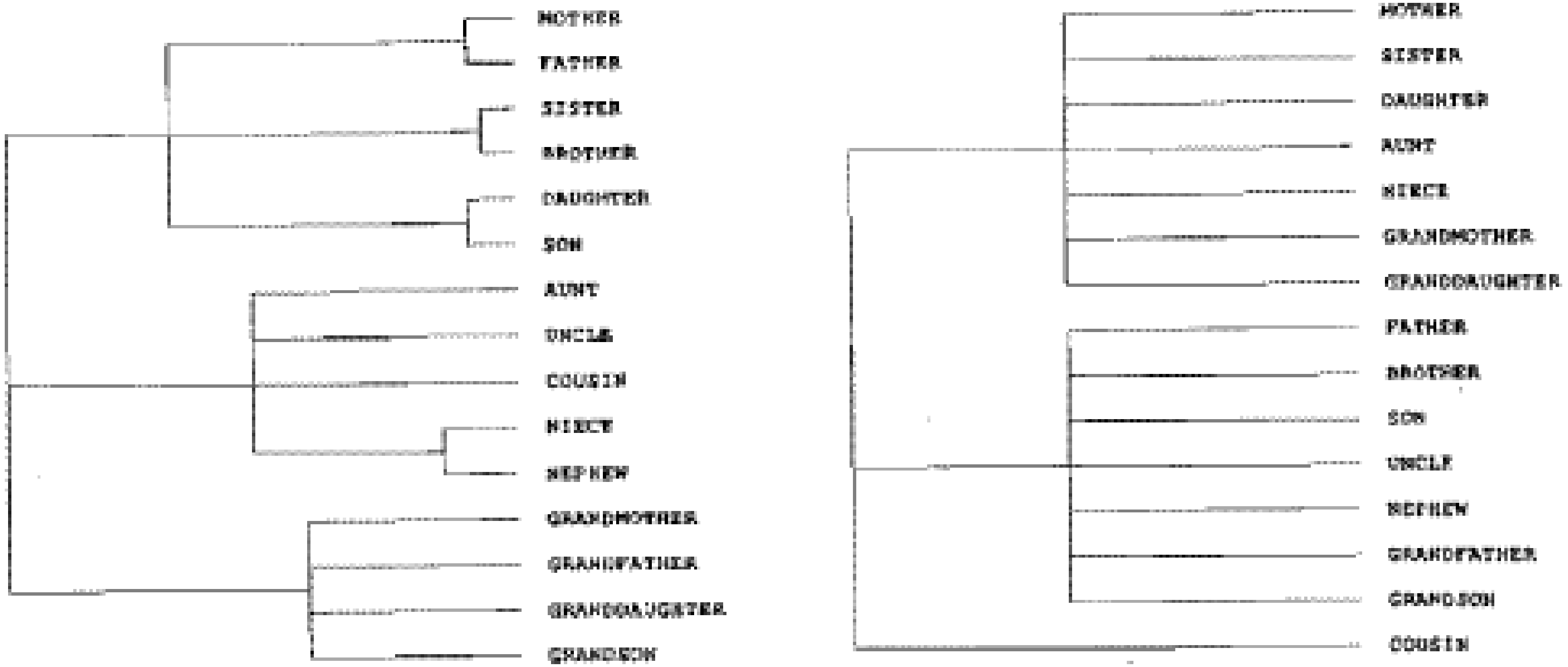
TABLE 2

Nine-Cluster MAPCLUS Solution for 15 Kinship Terms.

(1)	.582	BROTHR SISTER
(2)	.554	FATHER MOTHER
(3)	.551	DAUGHT SON
(4)	.547	GRDAUG GRFATH GRMOTH GRSON
(5)	.460	AUNT COUSIN NEPHEW NIECE
(6)	.432	NEPHEW NIECE
(7)	.398	AUNT DAUGHT GRDAUG GRMOTH MOTHER NIECE SISTER
(8)	.395	BROTHR FATHER GRFATH GRSON NEPHEW SON UNCLE
(9)	.311	BROTHR DAUGHT FATHER MOTHER SISTER SON

Representing overlapping clustering by multiple trees (cont.)

Two-tree representation for the MAPCLUS solution



A related problem: (Additive) clustering of 2-way 2-mode (rectangular proximity) data

(="block clustering", "bicluster" methods)

- Idea is to *simultaneously* cluster the rows and columns of a proximity matrix (e.g. consumers and products; people and clubs)
- Early work by Hartigan (1972; 1975)
- Recent papers by van Mechelen et al. (2004), Depril et al. (2008), Wilderjans et al. (2012)
- The clustering of rows and/or or columns can be overlapping (or not)

References – Nonhierarchical Clustering

Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review, 86(2),* 87-123.

Arabie, P., & Carroll, J. D. (1980). MAPCLUS: a mathematical programming approach to fitting the ADCLUS model. *Psychometrika, 45-2,* 211-235.

Carroll, J. D., & Arabie, P. (1983). INDCLUS: An individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika*, *48*(2), 157-169.

Ten Berge, J. M. F., & Kiers, H. A. L. (2005). A Comparison of Two Methods for Fitting the INDCLUS Model. *Journal of Classification, 22(2),* 273-286.

Arabie, P., Carroll, J. D., & DeSarbo, W. S. (1987). *Three-Way Scaling and Clustering.* Newbury Park: Sage.

Van Mechelen, I., Bock, H. H., & De Boeck, P. (2004). Two-mode clustering methods: a structured overview. *Statistical Methods in Medical Research,13(5), 363-394.*

*Depril, D., Van Mechelen, I., & Mirkin, B. (2008). Algorithms for additive clustering of rectangular data tables. *Computational Statistics and Data Analysis, 52,* 4923–4938.

Wilderjans, T. F., Depril, D., & Van Mechelen, I. (2012). Block-relaxation approaches for fitting the INDCLUS model. *Journal of Classification*, *29*(3), 277-296.

Corter, J.E., & Tversky, A. (1986). Extended similarity trees. *Psychometrika, 51,* 429-451.