Introduction to Multidimensional Scaling, Clustering, and Related Methods

James E. Corter

Session 1: Introduction and Background

MDSCN

"*MDSCN*" = MultiDimensional Scaling, Clustering, and Network models

→We will discuss methods to fit these three general classes of models (spaces, discrete-feature models, networks) to represent proximity relations among entities

-MDSCN methods are usually applied to *proximity* data, i.e. to the similarities, associations, or distances between each pair of entities

-MDSCN methods / models can also be applied to *multivariate data* (usually via the preliminary step of calculating a proximity measure).

-Some MDSCN methods have also been developed for *dominance or preference data*

Examples of proximity data

→ similarities or distances among two objects

- *Travel times* among cities
- *Similarity ratings* of pairs of concepts, or consumer products
- *Confusions* among stimuli (same-different; identification confusions)
- *Social distance* (e.g., communication frequency)

(Q: can correlations be considered as proximities?)

Example proximity data: similarity ratings of colors

A

Judged similarities between colors (Ekman)

Wavelength (nanometers)	434	445	465	472	490	504	537	555	584	600	610	628	651	674
434														
445	.86													
465	.42	.50												
472	.42	.44	.81											
490	.18	.22	.47	.54										
504	.06	.09	.17	.25	.61									
537	.07	.07	.10	.10	.31	.62								
555	.04	.07	.08	.09	.26	.45	.73							
584	.02	.02	.02	.02	.07	.14	.22	.33						
600	.07	.04	.01	.01	.02	.08	.14	.19	.58					
610	.09	.07	.02	.00	.02	.02	.05	.04	.37	.74				
628	.12	.11	.01	.01	.01	.02	.02	.03	.27	.50	.76			
651	.13	.13	.05	.02	.02	.02	.02	.02	.20	.41	.62	.85		
674	.16	.14	.03	.04	.00	.01	.00	.02	.23	.28	.55	.68	.76	

Wavelength (nanometers)

(ref. Shepard 1980)

Rated dissimilarity of 15 emotion words (Corter, 1980)

	happy	calm	elated	amused	worried	angry	depress	excited	frightn.	irritated	anxious	bored	sad	contnt.	hopeful
happy	--														
calm	4.05	--													
elated	2.44	7.83	--												
amused	3.35	6.52	4.81	--											
worried	9.52	9.19	9.23	8.63	--										
angry	10.21	10.23	9.93	9.28	5.87	--									
depressed	10.09	6.33	10.39	9.04	3.06	6.38	--								
excited	3.33	10.07	2.76	5.27	6.75	5.75	10.14	--							
frightened	9.17	9.81	8.84	8.82	2.86	5.00	6.18	4.55	--						
irritated	9.57	9.46	9.25	7.73	5.16	2.38	5.53	5.64	6.21	--					
anxious	7.00	9.64	7.19	7.94	2.73	5.87	6.11	3.57	4.73	5.66	--				
bored	8.78	5.19	9.05	9.68	7.43	7.38	5.31	9.71	9.29	6.47	6.73	--			
sad	10.64	6.29	9.94	8.31	3.50	6.00	2.42	9.12	5.70	6.42	6.13	4.80	--		
contented	3.70	2.43	4.80	4.12	9.52	9.39	9.27	5.50	9.60	9.66	9.00	8.66	8.92	--	
hopeful	5.13	5.27	5.26	5.68	6.83	9.20	9.11	2.59	8.35	8.64	6.27	8.37	7.46	5.94	--

(E. Rothkopf’s Morse code data -- same-different confusions)

D

Signal presented first (rows) × Signal presented second (columns)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0	
A	92	04	06	13	03	14	10	13	46	05	22	03	25	34	06	06	09	35	23	06	37	13	17	12	07	03	02	07	05	05	08	06	05	06	02	03	A
B	05	84	37	31	05	28	17	21	05	19	34	40	06	10	12	22	25	16	18	02	18	34	08	84	30	42	12	17	14	40	32	74	43	17	04	04	B
C	04	38	87	17	04	29	13	07	11	19	24	35	14	03	09	51	34	24	14	06	06	11	14	32	82	38	13	15	31	14	10	30	28	24	18	12	C
D	08	62	17	88	07	23	40	36	09	13	81	56	08	07	09	27	09	45	29	06	17	20	27	40	15	33	03	09	06	11	09	19	08	10	05	06	D
E	06	13	14	06	97	02	04	04	17	01	05	06	04	04	05	01	05	10	07	67	03	03	02	05	06	05	04	03	05	03	05	02	04	02	03	03	E
F	04	51	33	19	02	90	10	29	05	33	16	50	07	06	10	42	12	35	14	02	21	27	25	19	27	13	08	16	47	25	26	24	21	05	05	05	F
G	09	18	27	38	01	14	90	06	05	22	33	16	14	13	82	52	23	21	05	03	15	14	32	21	23	39	15	14	05	10	04	10	17	23	20	11	G
H	03	45	23	25	09	32	08	87	10	10	09	29	05	08	08	14	08	17	37	04	36	59	09	33	14	11	03	09	15	43	70	35	17	04	03	03	H
I	64	07	07	13	10	08	06	12	93	03	05	16	13	30	07	03	05	19	35	16	10	05	08	02	05	07	02	05	08	09	06	08	05	02	04	05	I
J	07	09	38	09	02	24	18	05	04	85	22	31	08	03	21	63	47	11	02	07	09	09	09	22	32	28	67	66	33	15	07	11	28	29	26	23	J
K	05	24	38	73	01	17	25	11	05	27	91	33	10	12	31	14	31	22	02	02	23	17	33	63	16	18	05	09	17	08	08	18	14	13	05	06	K
L	02	69	43	45	10	24	12	26	09	30	27	86	06	02	09	37	36	28	12	05	16	19	20	31	25	59	12	13	17	15	26	29	36	16	07	03	L
M	24	12	05	14	07	17	29	08	08	11	23	08	96	62	11	10	15	20	07	09	13	04	21	09	18	08	05	07	06	06	05	07	11	07	10	04	M
N	31	04	13	30	08	12	10	16	13	03	16	08	59	93	05	09	05	28	12	10	16	04	12	04	06	11	05	02	03	04	04	06	02	02	10	02	N
O	07	07	20	06	05	09	76	07	02	39	26	10	04	08	86	37	35	10	03	04	11	14	25	35	27	27	19	17	07	07	06	18	14	11	20	12	O
P	05	22	33	12	05	36	22	12	03	78	14	46	05	06	21	83	43	23	09	04	12	19	19	19	41	30	34	44	24	11	15	17	24	23	25	13	P
Q	08	20	38	11	04	15	10	05	02	27	23	26	07	06	22	51	91	11	02	03	06	14	12	37	50	63	34	32	17	12	09	27	40	58	37	24	Q
R	13	14	16	23	05	34	26	15	07	12	21	37	14	12	12	29	08	87	16	02	23	23	62	14	12	13	07	10	13	04	07	12	07	09	01	02	R
S	17	24	05	30	11	26	05	59	16	03	13	10	05	17	06	06	03	18	96	09	56	24	12	10	06	07	08	02	02	15	28	09	05	05	05	02	S
T	13	10	01	05	46	03	06	06	14	06	14	07	06	05	06	11	04	04	07	96	08	05	04	02	02	06	05	05	03	03	03	08	07	06	14	06	T
U	14	29	12	32	04	32	11	34	21	07	44	32	11	13	06	20	12	40	51	06	93	57	34	17	09	11	06	06	16	34	10	09	09	07	04	03	U
V	05	17	24	16	09	29	06	39	05	11	26	43	04	01	09	17	10	17	11	06	32	92	17	57	35	10	10	14	28	79	44	36	25	10	01	05	V
W	09	21	30	22	09	36	25	15	04	25	29	18	15	06	26	20	25	61	12	04	19	20	86	22	25	22	10	22	19	16	05	09	11	06	03	07	W
X	07	64	45	19	03	28	11	06	01	35	50	42	10	08	24	32	61	10	12	03	12	17	21	91	48	26	12	20	24	27	16	57	29	16	17	06	X
Y	09	23	62	15	04	26	22	09	01	30	12	14	05	06	14	30	52	05	07	04	06	13	21	44	86	23	26	44	40	15	11	26	22	33	23	16	Y
Z	03	46	45	18	02	22	17	10	07	23	21	51	11	02	15	59	72	14	04	03	09	11	12	36	42	87	16	21	27	09	10	25	66	47	15	15	Z
1	02	05	10	03	03	05	13	04	02	29	05	14	09	07	14	30	28	09	04	02	03	12	14	17	19	22	84	63	13	08	10	08	19	32	57	55	1
2	07	14	22	05	04	20	13	03	25	26	09	14	02	03	17	37	28	06	05	03	06	10	11	17	30	13	62	89	54	20	05	14	20	21	16	11	2
3	03	08	21	05	04	32	06	12	02	23	06	13	05	02	05	37	19	09	07	06	04	16	06	22	25	12	18	64	86	31	23	41	16	17	08	10	3
4	06	19	19	12	06	25	14	16	07	21	13	19	03	03	02	17	29	11	09	03	17	55	08	37	24	03	05	26	44	89	42	44	32	10	03	03	4
5	08	45	15	14	02	45	04	67	07	14	04	41	02	00	04	13	07	09	27	02	14	45	07	45	10	10	14	10	30	69	90	42	24	10	06	05	5
6	07	80	30	17	04	23	04	14	02	11	11	27	06	02	07	16	30	11	14	03	12	30	09	58	38	39	15	14	26	24	17	86	69	14	05	14	6
7	06	33	22	14	05	25	06	04	06	24	13	32	07	06	07	36	39	12	06	02	03	13	09	30	30	50	22	29	18	15	12	61	85	70	20	13	7
8	03	23	40	06	03	15	15	06	02	33	10	14	03	06	14	12	45	02	06	04	06	07	05	24	35	50	42	29	16	16	09	30	60	89	61	26	8
9	03	14	23	03	01	06	14	05	02	30	06	07	16	11	10	31	32	05	06	07	06	03	08	11	21	24	57	39	09	12	04	11	42	56	91	78	9
0	09	03	11	02	05	07	14	04	05	30	08	03	02	03	25	21	29	02	03	04	05	03	02	12	15	20	50	26	09	11	05	22	17	52	81	94	0
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	1	2	3	4	5	6	7	8	9	0	

Signal presented second

A quick example of (<u>metric</u>) MDS:

Data: a matrix of dissimilarities (here, the dissimilarities are the physical distances among a sample of US cities) (example from Kruskal & Wish, 1970)
So in this example, the question is: can we reconstruct the two-dimensional map of the US from these interpoint distances?

Table 1 Flying Mileages Between 10 American Cities

Atlanta	Chicago	Denver	Houston	Los Angeles	Miami	New York	San Francisco	Seattle	Washington, DC	
0	587	1212	701	1936	604	748	2139	2182	543	Atlanta
587	0	920	940	1745	1188	713	1858	1737	597	Chicago
1212	920	0	879	831	1726	1631	949	1021	1494	Denver
701	940	879	0	1374	968	1420	1645	1891	1220	Houston
1936	1745	831	1374	0	2339	2451	347	959	2300	Los Angeles
604	1188	1726	968	2339	0	1092	2594	2734	923	Miami
748	713	1631	1420	2451	1092	0	2571	2408	205	New York
2139	1858	949	1645	347	2594	2571	0	678	2442	San Francisco
2182	1737	1021	1891	959	2734	2408	678	0	2329	Seattle
543	597	1494	1220	2300	923	205	2442	2329	0	Washington, DC

The MDS statistical problem: <u>try to find a configuration of n points in R-dimensional space such that the interpoint distances perfectly model the data "dissimilarities"</u> → here, find configuration matrix X = coordinates of 10 points in 2-dimensional Euclidean space that recreate the data dissimilarities

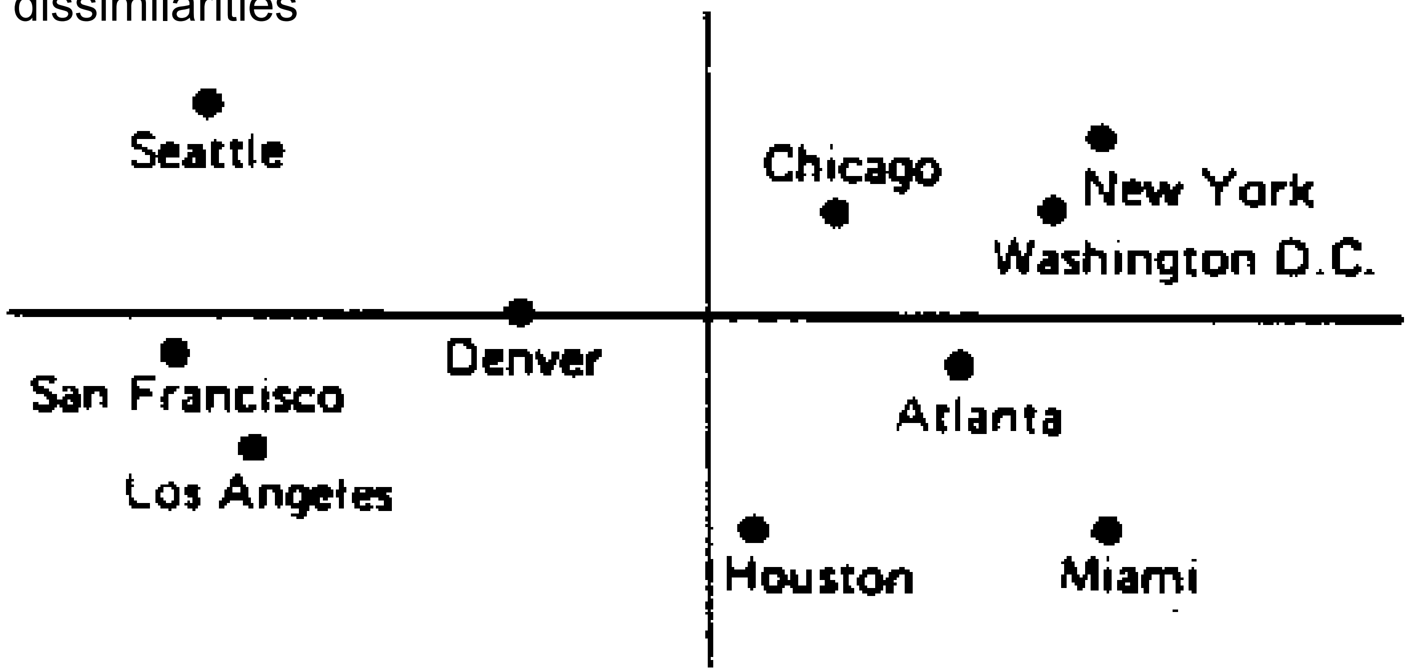


Figure 1 CMDS of flying mileages between 10 American cities.

For "metric MDS": If the solution is optimal (with perfect fit), then "model distances" in the derived map space should be <u>linearly related</u> to the dissimilarity data (here, airline distances). For this example, they will in fact be directly proportional).

Example 2: “nonmetric” MDS analysis from Shepard (1980)

A

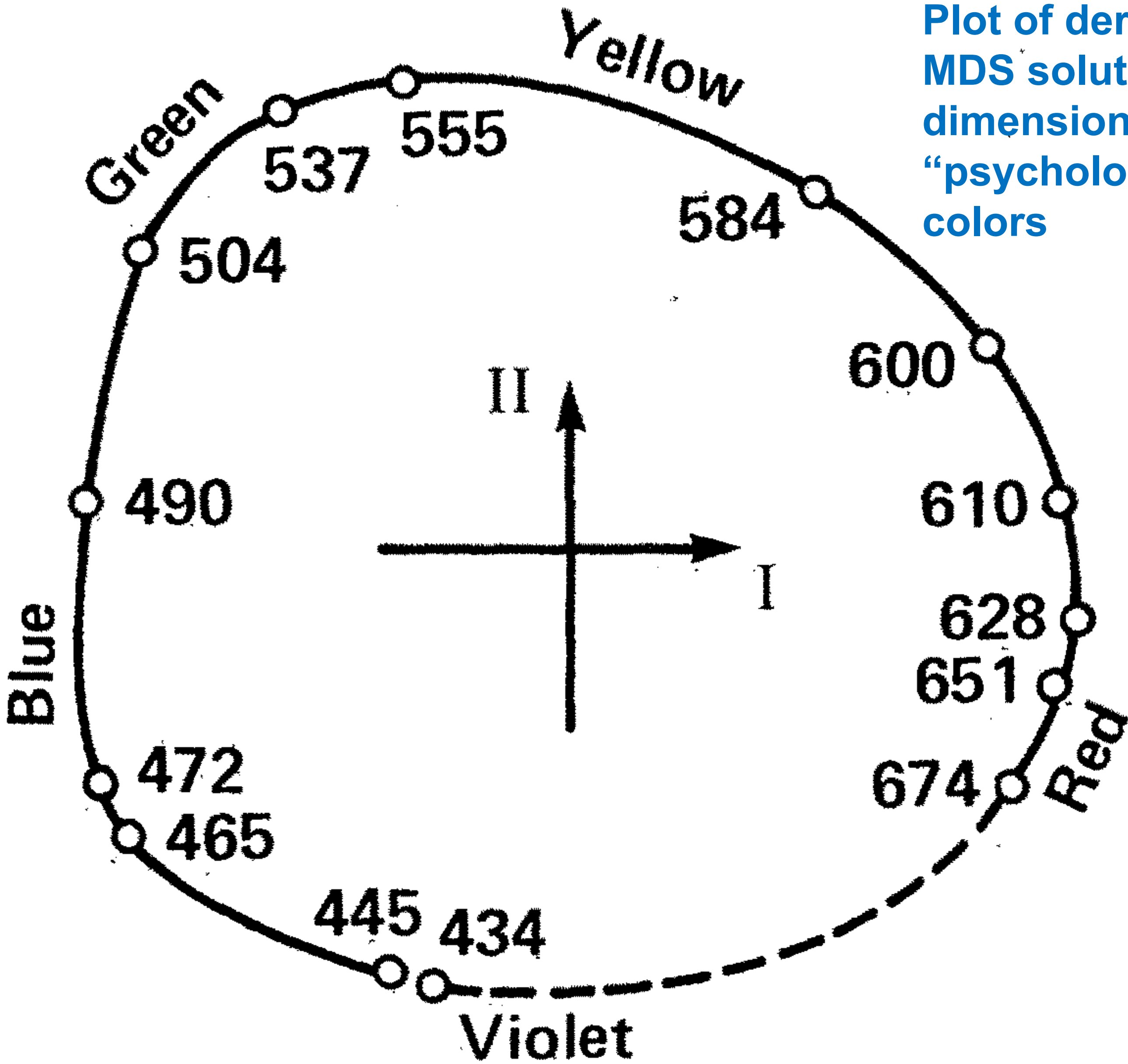
Judged similarities between colors (Ekman)

Wavelength (nanometers)

	434	445	465	472	490	504	537	555	584	600	610	628	651	674
434														
445	.86													
465	.42	.50												
472	.42	.44	.81											
490	.18	.22	.47	.54										
504	.06	.09	.17	.25	.61									
537	.07	.07	.10	.10	.31	.62								
555	.04	.07	.08	.09	.26	.45	.73							
584	.02	.02	.02	.02	.07	.14	.22	.33						
600	.07	.04	.01	.01	.02	.08	.14	.19	.58					
610	.09	.07	.02	.00	.02	.02	.05	.04	.37	.74				
628	.12	.11	.01	.01	.01	.02	.02	.03	.27	.50	.76			
651	.13	.13	.05	.02	.02	.02	.02	.02	.20	.41	.62	.85		
674	.16	.14	.03	.04	.00	.01	.00	.02	.23	.28	.55	.68	.76	

Wavelength (nanometers)

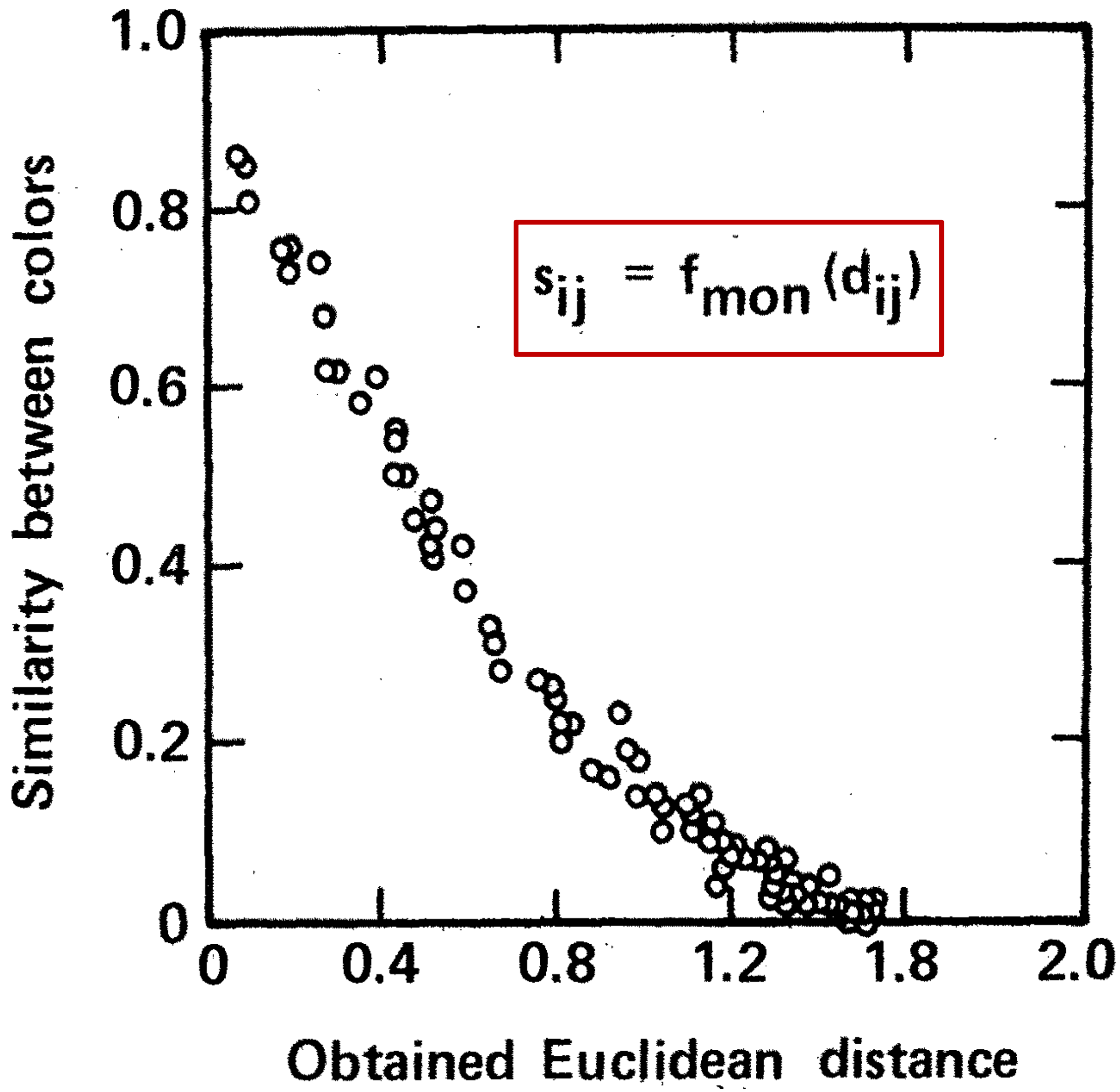
(ref. Shepard 1980)



Plot of derived nonmetric MDS solution in R=2 dimensions → the "psychological space" of colors

(ref. Shepard 1980)

nonmetric MDS: distances in space (d) only monotonically related to proximities (s)



(ref. Shepard 1980)

Some uses of MDSCN methods:

1. Exploratory data analysis

1.5 ”Unsupervised learning” technique

2. Mathematical modeling of (the proximities resulting from) some specific cognitive, behavioral, or other real-world process

If we are doing mathematical modeling, then:

-the model should be (theoretically) appropriate to the data, and vice-versa

-the model should be as parsimonious as possible, while still providing good fit

-ideally, the parameters of the model have interesting interpretations

Uses of MDSCN (cont.)

Borg & Groenen (2010) cite four major uses of MDS (or other "scaling" technique):

1. to explore structure in data
2. to test structural hypotheses
3. to explore psychological structure (of "mental spaces")
4. as a mathematical model of similarity judgments

Proximities can also be derived from **multivariate data** (observations x variables) – e.g. correlations can be computed from a multivariate matrix Y

TABLE 1.1. Correlations of crime rates over 50 U.S. states.

Crime	No.	1	2	3	4	5	6	7
Murder	1	1.00	0.52	0.34	0.81	0.28	0.06	0.11
Rape	2	0.52	1.00	0.55	0.70	0.68	0.60	0.44
Robbery	3	0.34	0.55	1.00	0.56	0.62	0.44	0.62
Assault	4	0.81	0.70	0.56	1.00	0.52	0.32	0.33
Burglary	5	0.28	0.68	0.62	0.52	1.00	0.80	0.70
Larceny	6	0.06	0.60	0.44	0.32	0.80	1.00	0.55
Auto theft	7	0.11	0.44	0.62	0.33	0.70	0.55	1.00

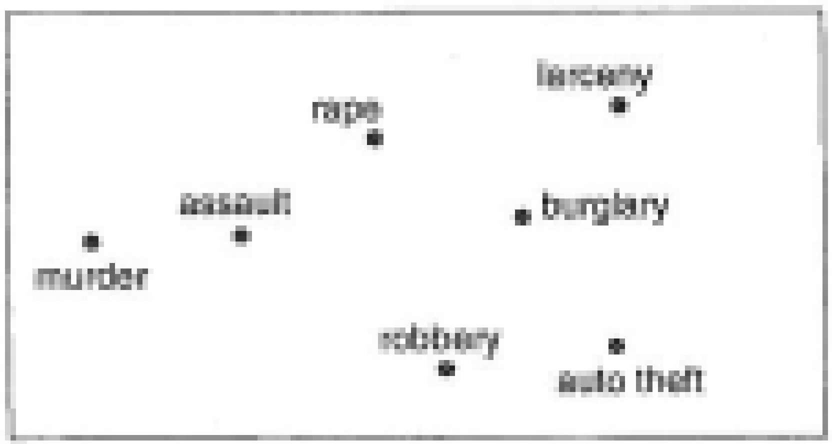


FIGURE 1.1. A two-dimensional MDS representation of the correlations in Table 1.1.

From Borg & Groenen (2005)

“Nonmetric” MDS as mathematical / psychometric modeling:
Analysis of Miller-Nicely data on consonant confusions (from Shepard 1980)

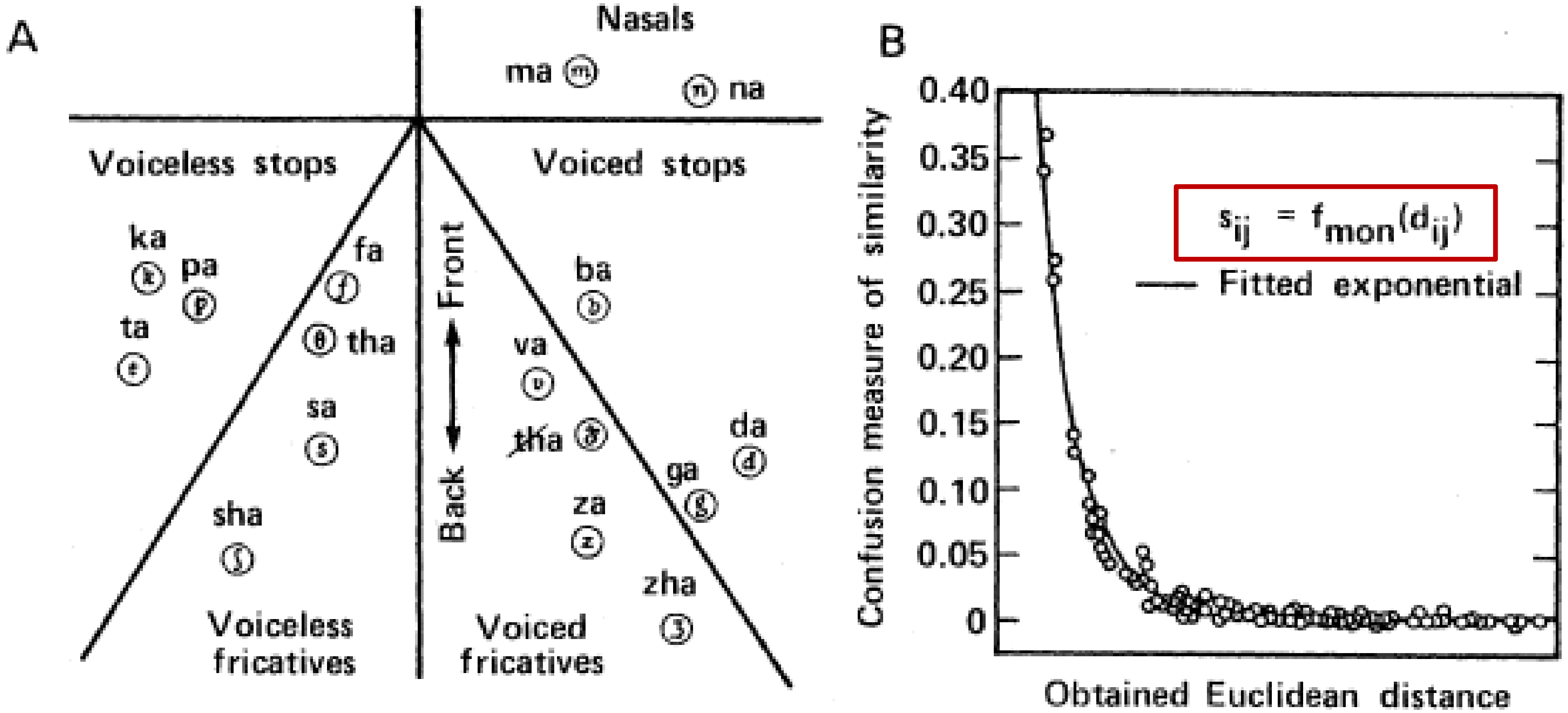


Fig. 3. Various multidimensional scalings of 16 English consonants based on the confusion data of Miller and Nicely (25). (A) the two-dimensional configuration of Fig. 2 with interpretive lines added. (B) Obtained (exponential) relation, for all pairs of the 16 consonants, between the confusion data and corresponding Euclidean distances in the obtained configuration in (A). [From

From Shepard 1980

Note that MDS-derived “maps” or configurations can be in 3 or more dimensions

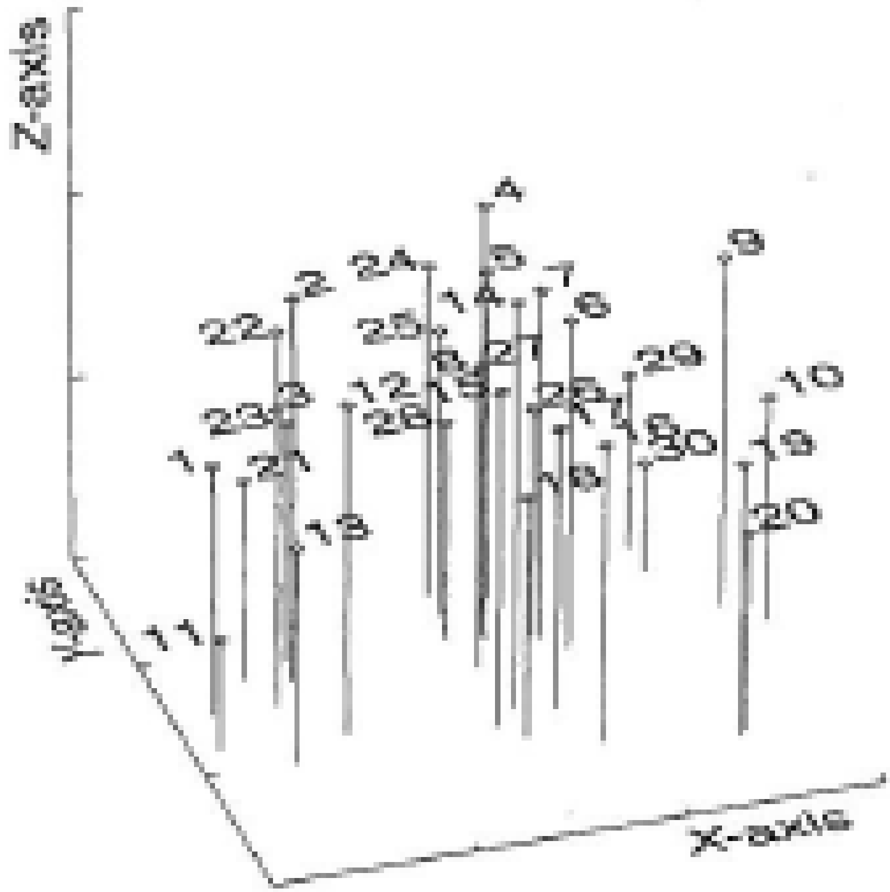
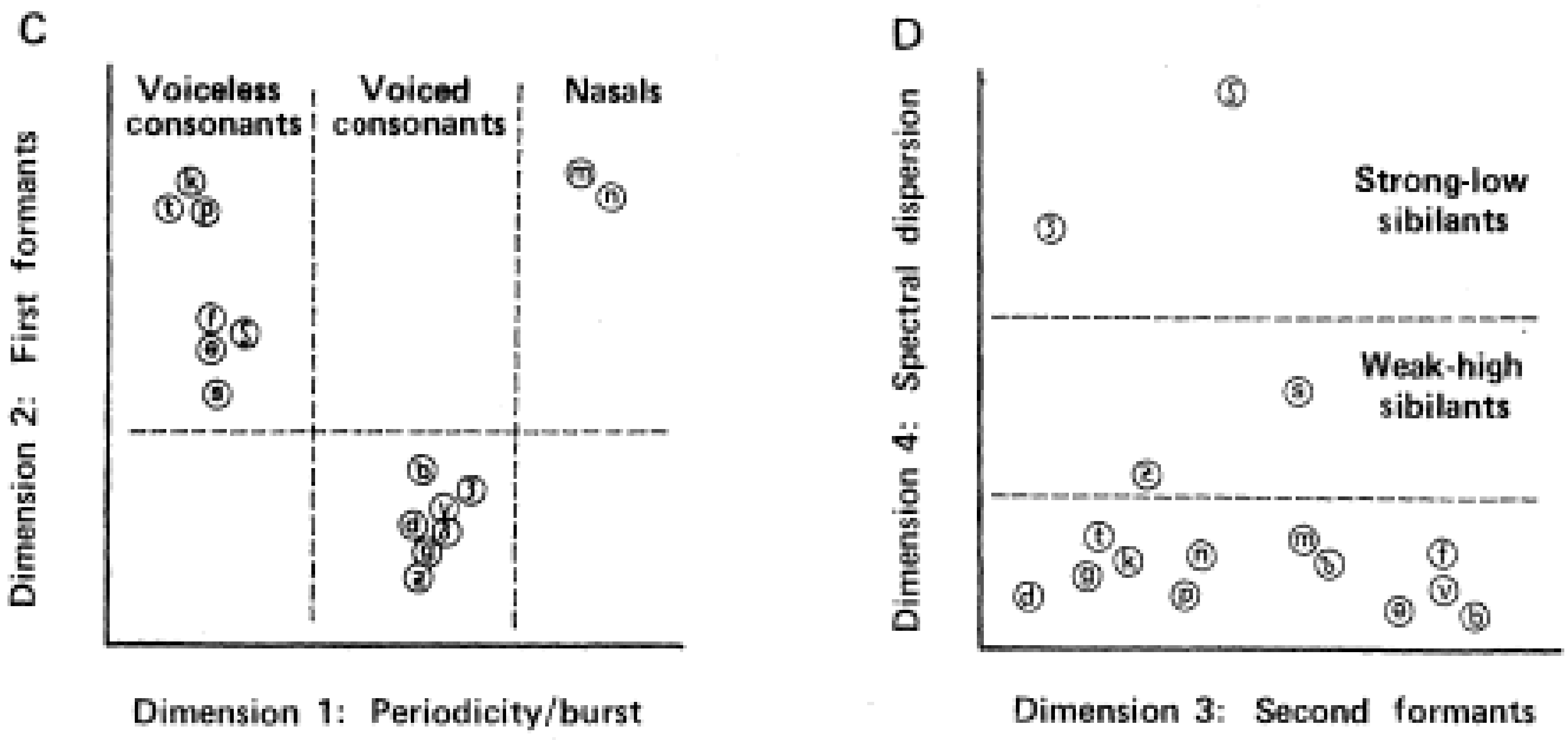


FIGURE 1.2. Three-dimensional MDS representation of protest acts.

From Shepard 1980

A 4-dimensional MDS solution for the Miller-Nicely confusions (from Shepard 1980)



From Shepard 1980

Four types of data for MDSCN scaling: proximity, preference, dominance, multivariate

1) ** Proximity data = defined on pairs of objects (sometimes between an object and an attribute) → "closeness" of entities

- -similarity ratings between all pairs of 14 color chips
- -Haming distance between two arrays
- -correlations among 20 test items

2) Preference = defined between people and a set of objects, products, etc.

- -preference rankings for ten job candidates by a committee
- -consumer rating of liking for 15 products (e.g., cell phone models)

3) Dominance data = defined on pairs of objects (or between an object and an attribute) → one entity is "superior to" or "beats" another

- -data on 45 students' performance on a 20-item test
- -binary pairwise choice data on a set of consumer products

4) Multivariate data = X = *n* objects measured on *m* variables (attributes)

- -50 patients measured on 14 physiological variables
- -college applicants with data on SAT scores, grades, etc.

NOTE: for any of these types, <u>measurement level</u> of the variables may vary

Similarities (from sorting task) on eight occupations

TreeModels

TABLE 1: An example proximity matrix. Data are similarities among 8 occupations, derived from a sorting task (Kraus, 1976).

	bio	phy	law	pol	ins	car	cab	con
biologist	---							
physician	812	---						
lawyer	727	714	---					
police officer	156	175	195	---				
insurance agent	123	156	162	221	---			
carpenter	39	65	45	110	227	---		
cab driver	52	78	39	149	188	396	---	
constr. worker	32	52	19	84	136	610	364	---

Data from Kraus, 1976 / in Corter, 1996

Example of rectangular preference data

TABLE 2: An example rectangular proximity data: preference rankings of 12 described jobs by 8 subjects.

	DESCRIBED JOB											
	j01	j02	j03	j04	j05	j06	j07	j08	j09	j10	j11	j12
SUB:	---	---	---	---	---	---	---	---	---	---	---	---
s01	4	1	12	7	2	3	8	6	5	11	9	10
s02	9	12	4	8	3	10	1	11	7	2	5	6
s03	11	2	1	7	6	8	3	12	9	5	4	10
s04	10	5	1	12	3	6	8	11	9	4	2	7
s05	7	4	2	1	8	12	5	10	6	9	3	11
s06	3	8	5	6	4	11	9	7	2	12	1	10
s07	7	4	6	11	5	8	1	10	3	2	12	9
s08	7	3	2	9	4	11	8	6	1	10	5	12

An example of dominance data: test performance

name	ID	Item1	Item2	Item3
Joe	s01	0	0	0
Mary	s02	0	0	1
Yuh-Jia	s03	1	0	1
Eric	s04	1	0	1
Desdemona	s05	1	0	1
Tomoko	s06	1	1	0
Jie	s07	1	1	1
Sara	s08	1	1	1

1 = correct, 0 = incorrect

An example of multivariate data

Table 1.1 City Crime

	Murder Manslaughter	Rape	Robbery	Assault	Burglary	Larceny	Auto Theft
Atlanta	16.5	24.8	106	147	1112	905	494
Boston	4.2	13.3	122	90	982	669	954
Chicago	11.6	24.7	340	242	808	609	645
Dallas	18.1	34.2	184	293	1668	901	602
Denver	6.9	41.5	173	191	1534	1368	780
Detroit	13.0	35.7	477	220	1566	1183	788
Hartford	2.5	8.8	68	103	1017	724	468
Honolulu	3.6	12.7	42	28	1457	1102	637
Houston	16.8	26.6	289	186	1509	787	697
Kansas City	10.8	43.2	255	226	1494	955	765
Los Angeles	9.7	51.8	286	355	1902	1386	862
New Orleans	10.3	39.7	266	283	1056	1036	776
New York	9.4	19.4	522	267	1674	1392	848
Portland	5.0	23.0	157	144	1530	1281	488
Tucson	5.1	22.9	85	148	1206	756	483
Washington	12.5	27.6	524	217	1496	1003	739

From the United States Statistical Abstract (1970) per 100,000 population.

Some Data Theory:

Level of measurement of a variable (scale types): (S. S. Stevens, 1946, 1951)

nominal = numbers are labels only

ordinal = observations can be ordered

interval = intervals can be ordered (& equated)

ratio = ratios of numbers are meaningful

(ratio scale = interval scale with a natural 0 point)

Data Theory (cont)

Levels of measurement and permissible transformations:

Assume we have a valid "scale" or variable measured at one of the following levels of measurement. Then: if X is a valid scale, then X' is also a valid scale, provided that:

Nominal scale: $X' = f(X)$, *f(.) is an isomorphism*

Ordinal scale: $X' = f(X)$, *f(.) is monotonic function*

Interval scale: $X' = aX + c$ *(linear or affine transformation)*

Ratio scale: $X' = aX$ (ratio transformation)

Carroll & Arabie's (1980) taxonomy of data types for scaling:

A. Number of modes
   1. One mode
   2. Two modes
   3. Three or more modes

B. Power of a given mode

   A mode's power is the number of times the mode is repeated in the N-way table.
   1. Monadic data (e.g. single stimulus data, as from an absolute judgment task). Power = 1
   2. Dyadic data (e.g. proximities data). Power = 2
      a. Symmetric
      b. Nonsymmetric
   3. Polyadic data (e.g. judgments of homogeneity of sets of three or more stimuli, or similarity of or preference for "portfolios" of a number of items from the same set). Power $\geqslant 3$

C. Number of ways, defined as total number of factors, whether repeated or not, defining the data array; N if table of data is N-way (exclusive of replications, which are not usually thought of as defining a separate mode or way unless there is a structure on the replications *and* the replications "mode" is explicitly included in the model).

Data taxonomy (cont.):

D. Scale type of data (after Stevens, but with some additions)
   1. Nominal
   2. Ordinal
   3. Interval
   4. Ratio (sometimes called "interval with rational origin")
   5. Positive ratio
   6. Absolute

E. Conditionality of data
   1. Unconditional data
   2. Row or column conditional data (Coombs 1964)
   3. Matrix conditional data
   4. Other types of conditional data

F. Completeness of data
   1. Complete data
   2. Incomplete data

G. Number and nature of replications
   1. Only one data *set* comprising the data array
   2. Two or more data sets
      a. Same scale type for each replication
      b. Different scale types for different replications

MDSCN methods: Three general classes of models for proximity (and other) data:

1. Geometric models (spaces)
2. Clustering (discrete-feature) models
3. Graphs / networks

(also: nondimensional scaling, multiple trees, extended trees, hybrid representations, etc.)

The Idea of Geometric Distance

A geometric (spatial) distance model is used to derive distances from the positions of points in the MDS solution.

1) geometric (Euclidean) distance between points *x* and *y*
   -in the plane: $d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$

   -in R-dimensional space: $d(x,y) = \sqrt{\sum_{r=1}^{R}(x_r - y_r)^2}$

2) other Minkowski metrics (in R-space):

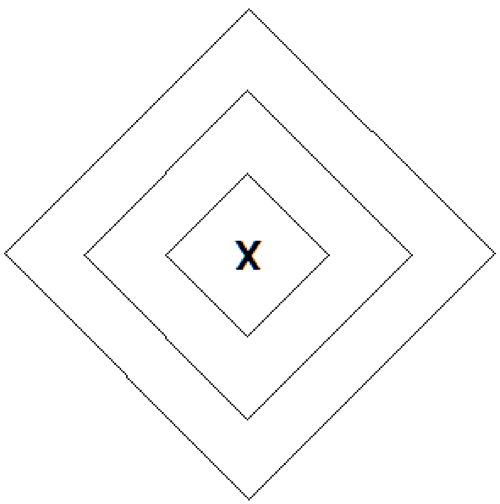
$$\mathrm{d}(x,y) = \left[\sum_{r=1}^{R} |x_r - y_r|^p\right]^{1/p} , p \geq 1$$

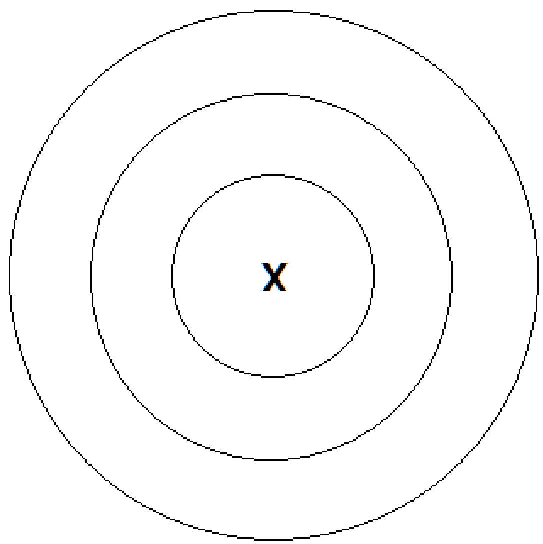
Minkowski metric: $$d_{xy} = \left[\sum_{r=1}^{R} \left| x_r - y_r \right|^p \right]^{1/p}, p \geq 1$$

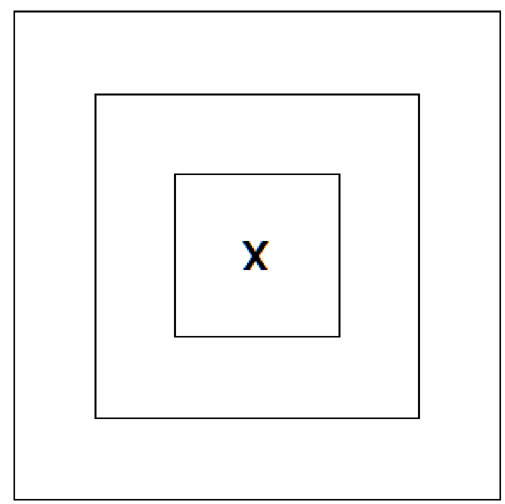
Three special cases:

p=1	p=2	p=infinity
"city-block" metric	Euclidean	"maximum" metric

Isosimilarity contours for these three metrics = the set(s) of all points that are equidistant from point X:







A more abstract definition of distance

(distances in a Euclidean or other geometric space satisfy these axioms, but so do other types of distances, e.g. the Haming distance)

→ *metric space* *(defined on a set of entities and the pairwise distances among them)*

Axioms of a metric space. Given a set of entities X, Y, Z,.. any distances defined among them must satisfy:

1) Symmetry: $d_{XY} = d_{YX}$
2) Nonnegativity: $d_{XY} \geq 0$
3) Minimality (nondegeneracy): $d_{XX} = 0$
   (i.e. $d_{XY} = 0$ if and only if $X = Y$)
4) Triangle inequality: $d_{XY} + d_{YZ} \geq d_{XZ}$

Note that this axiomatization is not minimal (Schreider, 1977) – axioms 1), 3) and 4) are sufficient to derive 2), for example.

Therefore, if we wish to represent a given set of (proximity?) data with an MDS (or other) model, we should ask:

Can the *proximity data* be considered as distances in a metric space? I.e., do the data satisfy the metric axioms?

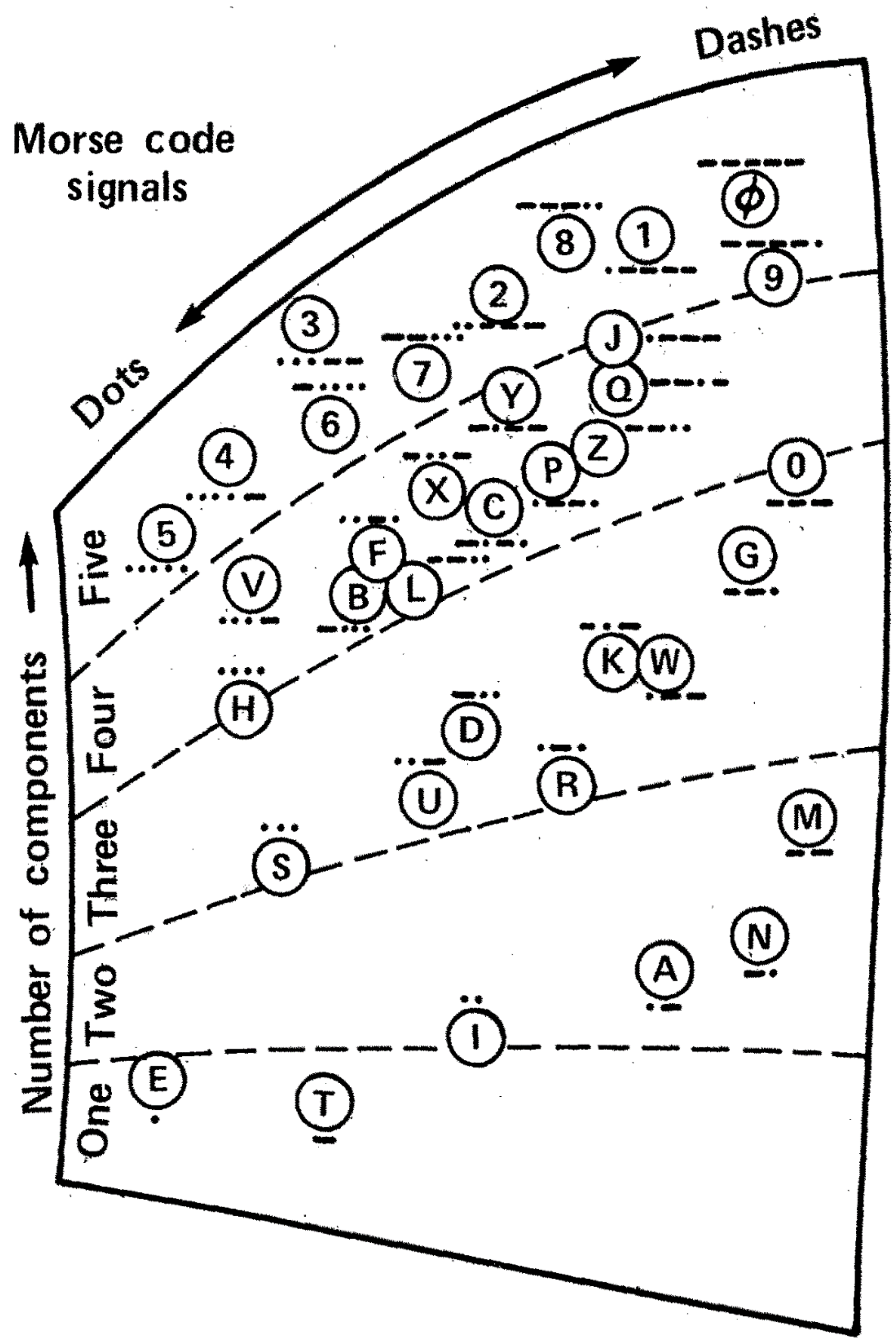
- If yes, then we can adopt geometric (or other metric distance) representation of these proximities
- If no, perhaps we can apply these distance models anyway (by first transforming data), or using a more general (less constrained) model relating data to model distances)

Three general classes of models for proximity (and other) data:

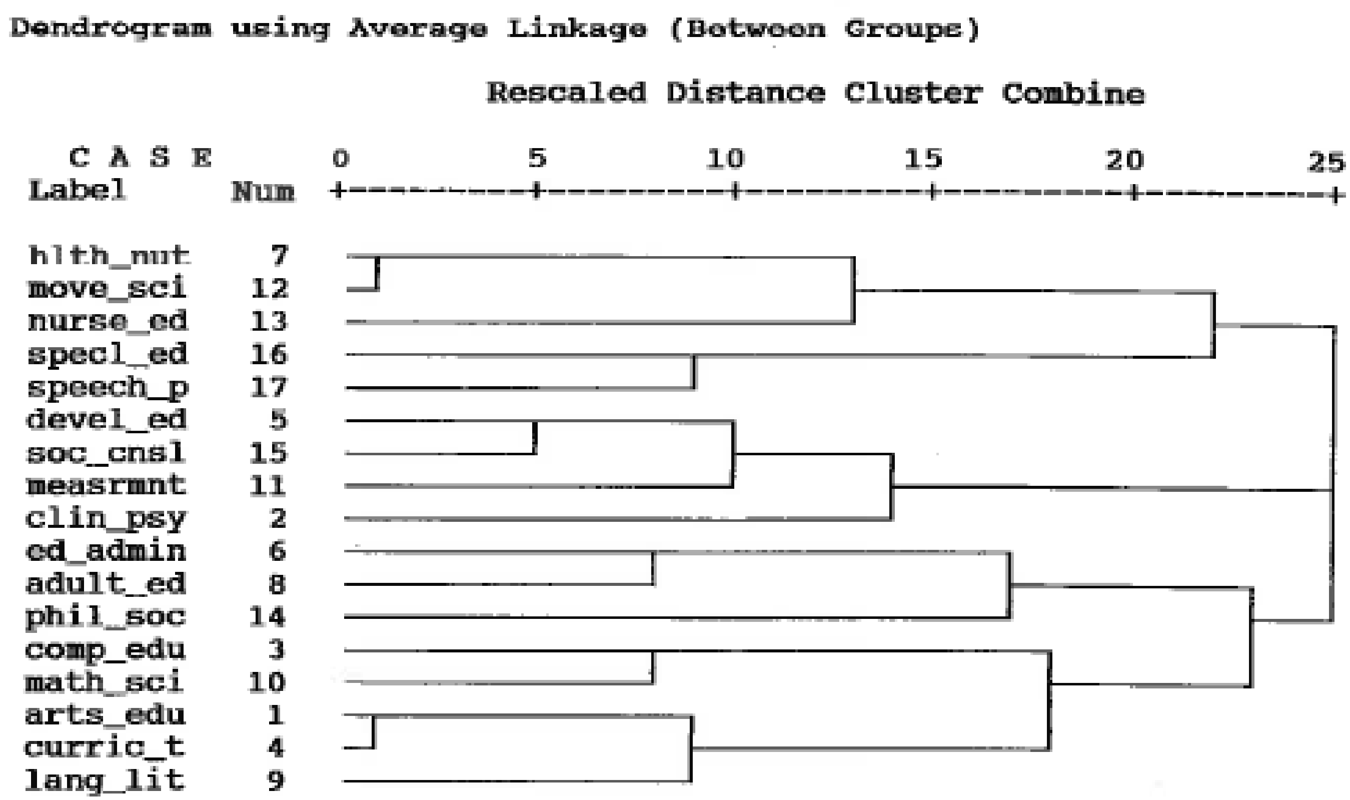
1. Geometric models (spaces)
2. Clustering (discrete-feature) models
3. Graphs / networks

EXAMPLES FOLLOW…

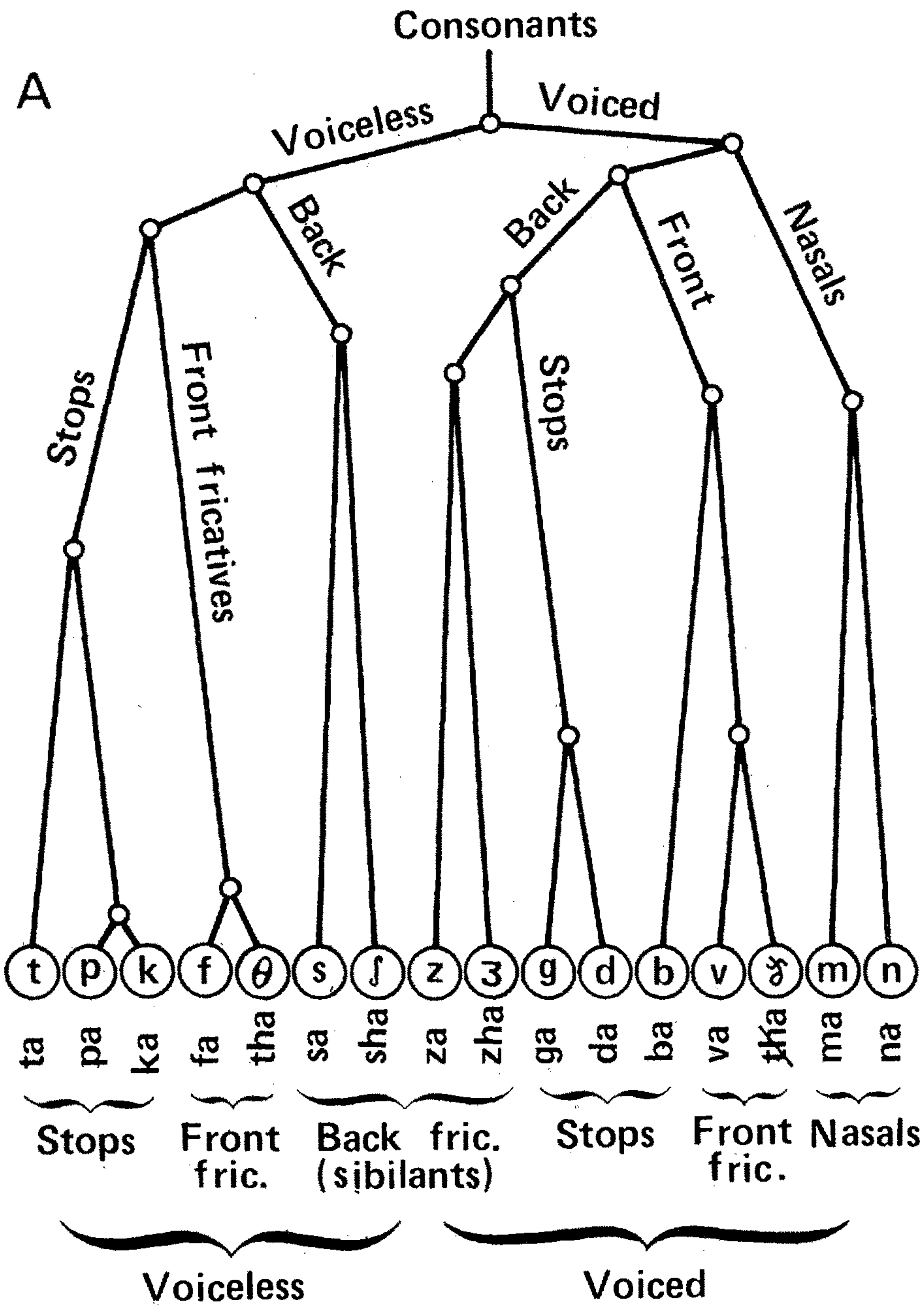
Rothkopf's Morse code data: 2-dim MDS solution



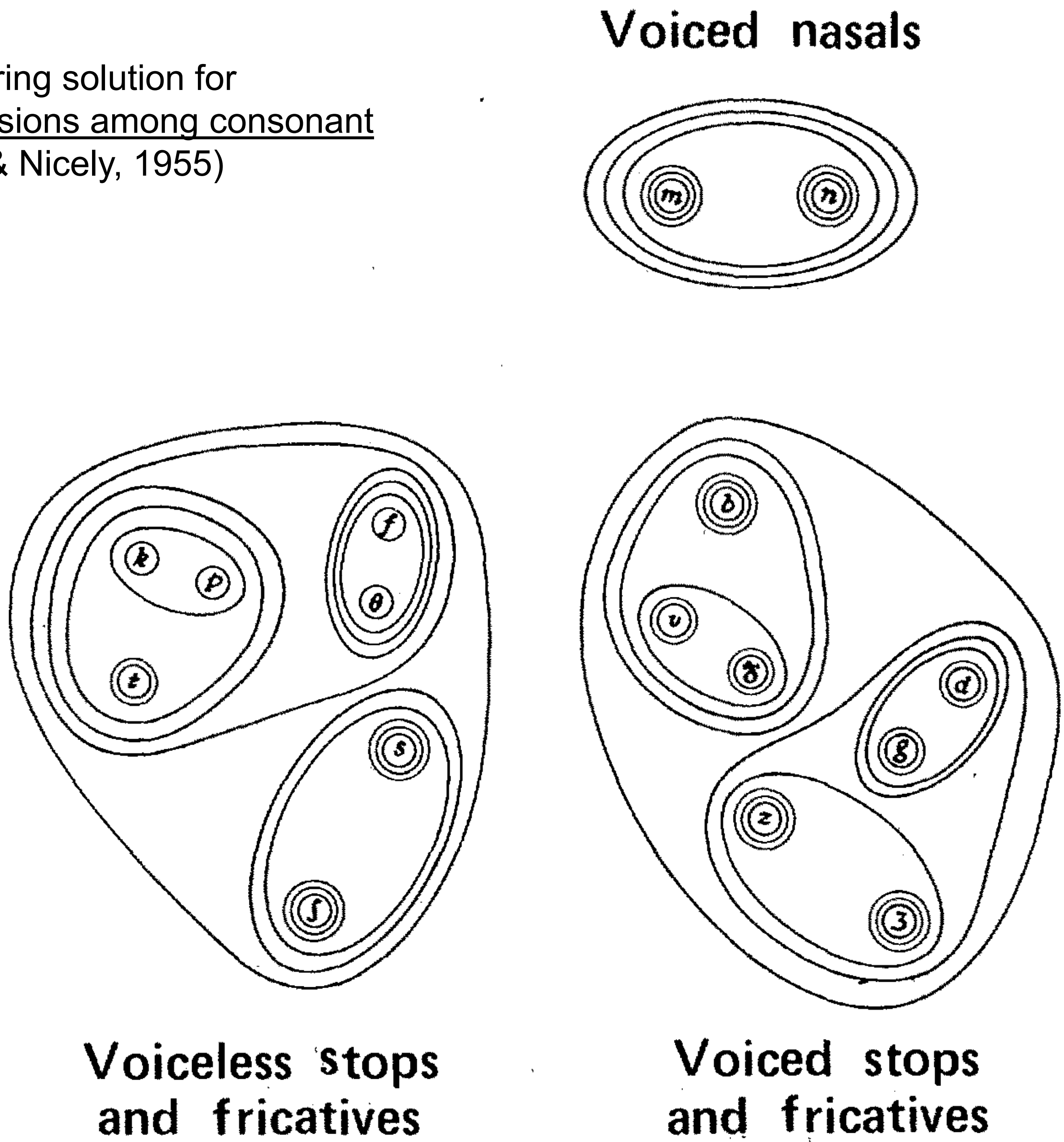
The dendrogram representing a hierarchical clustering can be viewed as an ultrametric tree (but fit is not optimized)



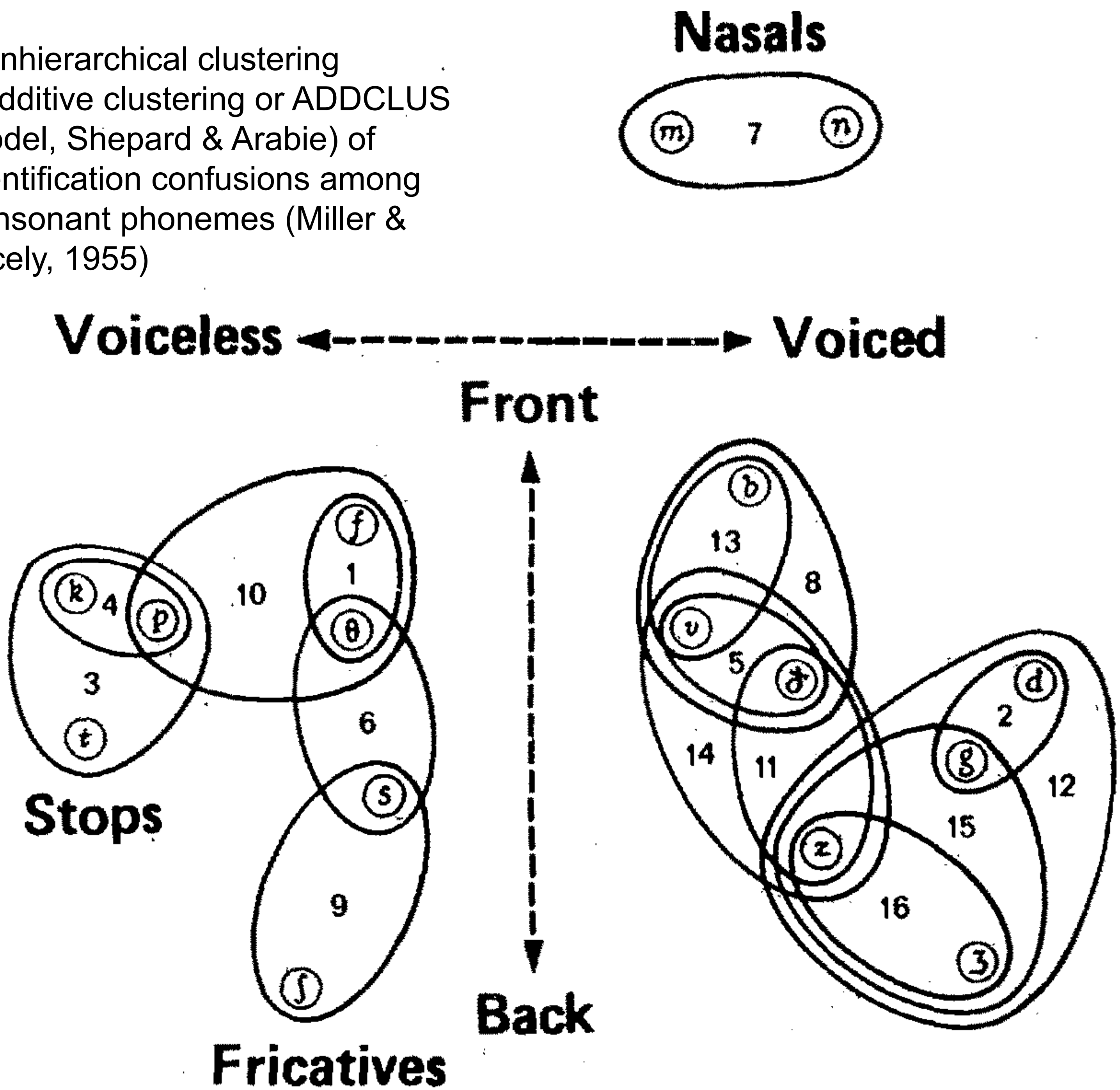
hierarchical clustering (ultrametric tree) solution for identification confusions among consonant phonemes (Miller & Nicely, 1955)



Hierarchical clustering solution for identification confusions among consonant phonemes (Miller & Nicely, 1955)



Nonhierarchical clustering ("additive clustering or ADDCLUS model, Shepard & Arabie) of identification confusions among consonant phonemes (Miller & Nicely, 1955)



Tree models: Below is shown an additive tree of hypothetical data on worker communication frequencies, inverted to dissimilarities by subtracting the similarities from a constant (25)

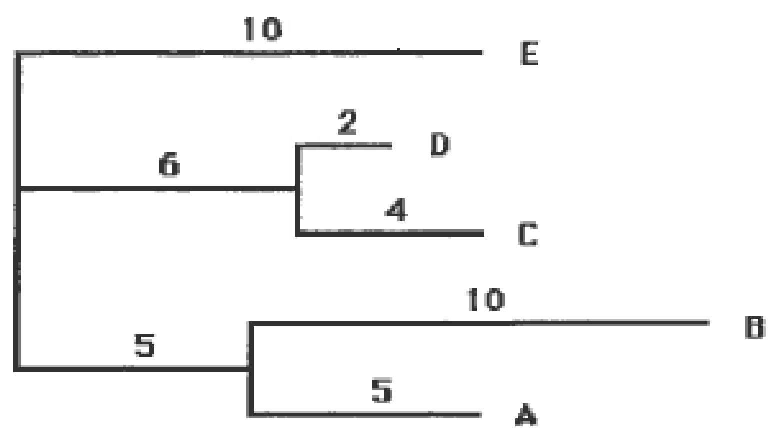
Data: communication frequencies among five coworkers

	A	B	C	D	E
Worker A	--				
Worker B	10	--			
Worker C	5	0	--		
Worker D	7	2	19	--	
Worker E	5	0	5	7	--

a. Dissimiliarities

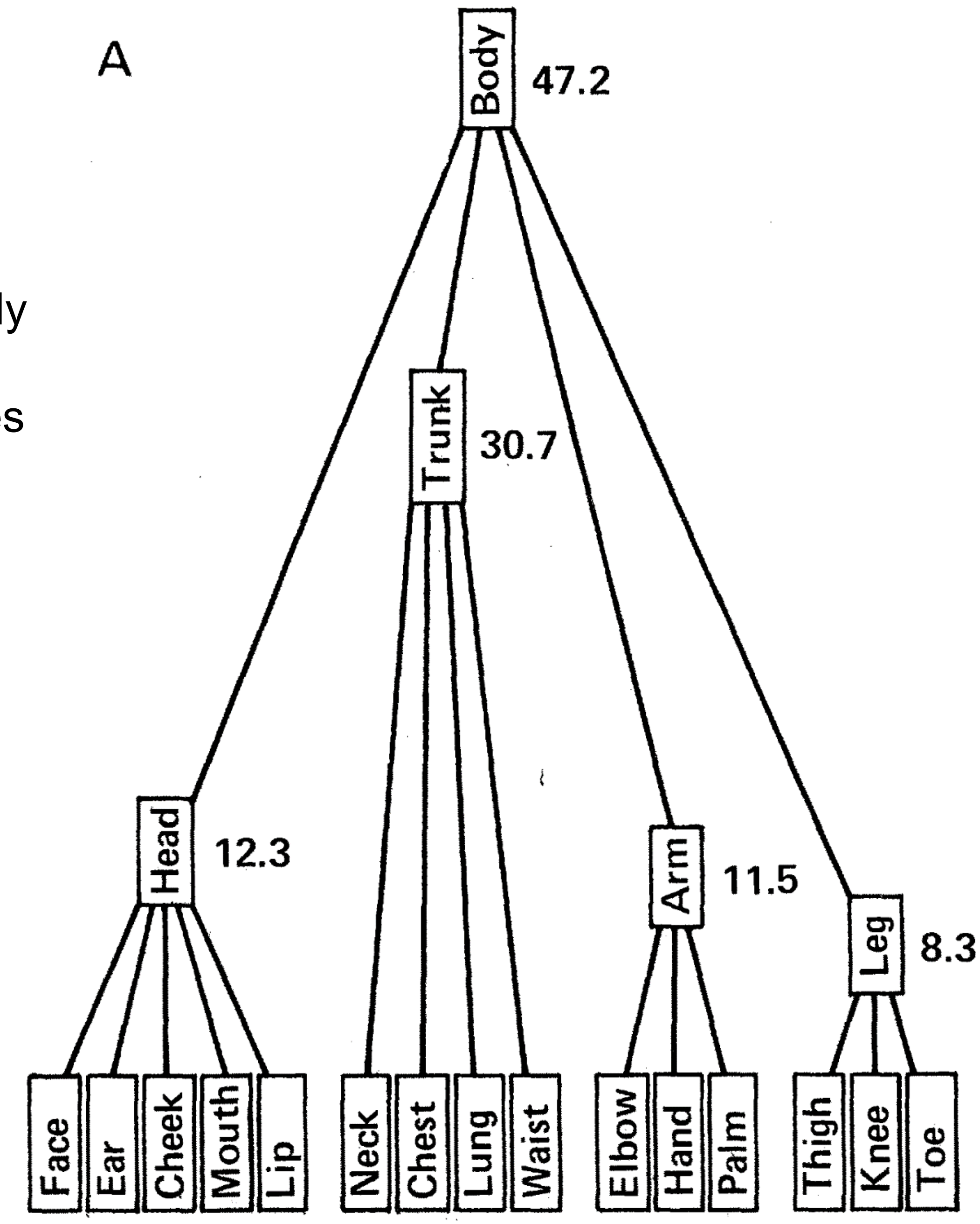
	A	B	C	D
Worker A	--			
Worker B	15	--		
Worker C	20	25	--	
Worker D	18	23	6	--
Worker E	20	25	20	18

b. Additive Tree



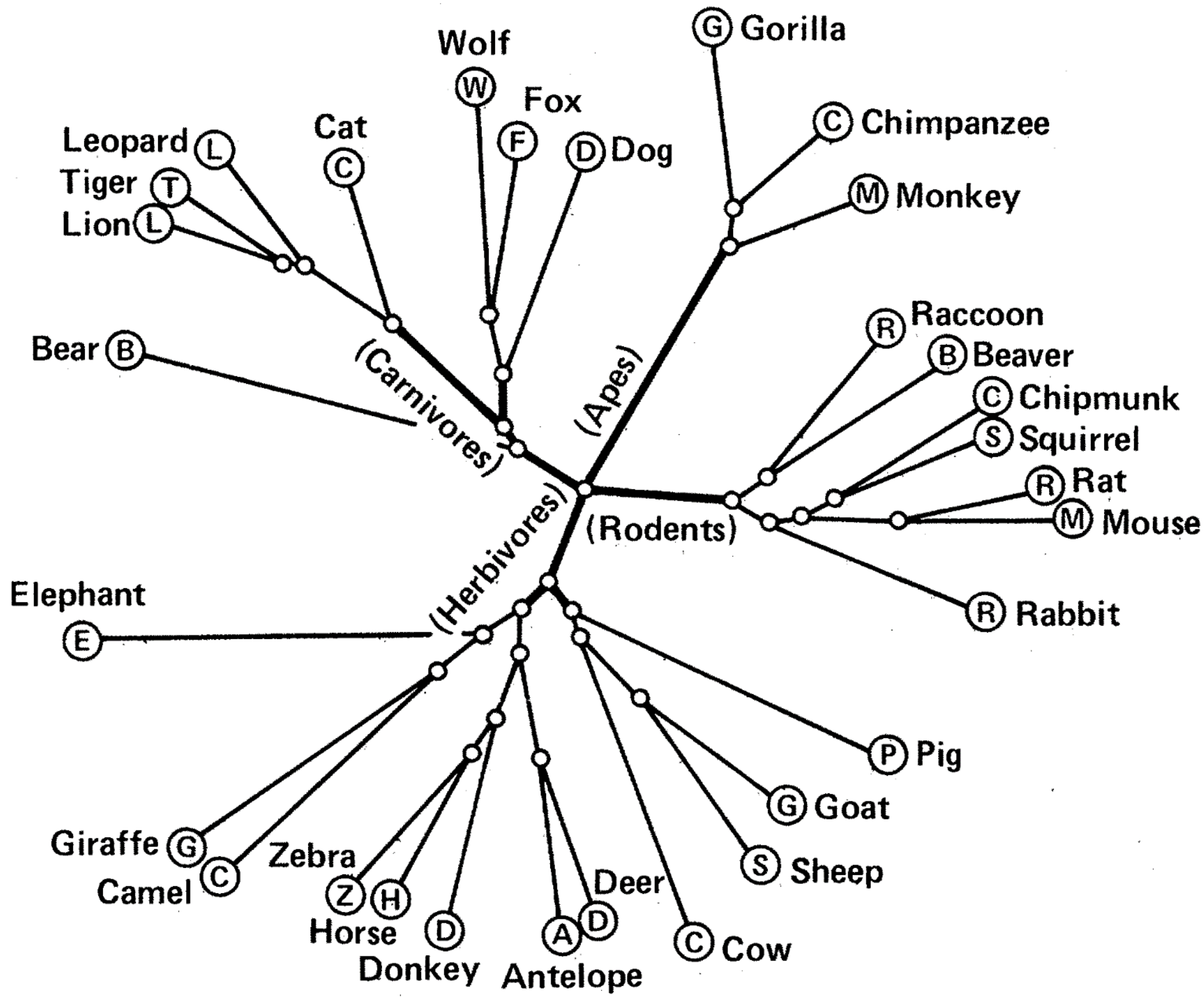
From Corter (1996)

A hierarchical network or tree model for semantic similarities among 20 body parts (from Miller) – note labeling of “internal” nodes



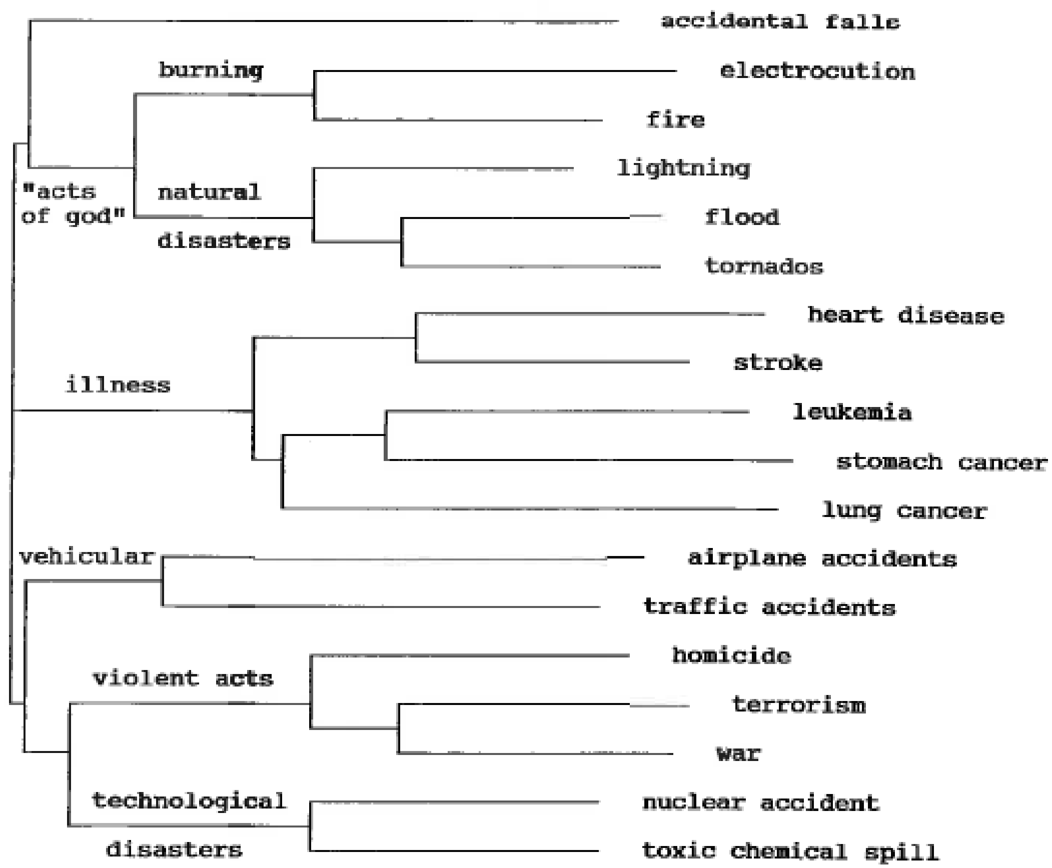
From Shepard (1980)

Fig. 5. Additive tree obtained by analysis of Henley's data on the conceptual similarities between 30 species of animal, embedded in a two-dimensional space. [Rearranged from Sattath and Tversky (*47*)]



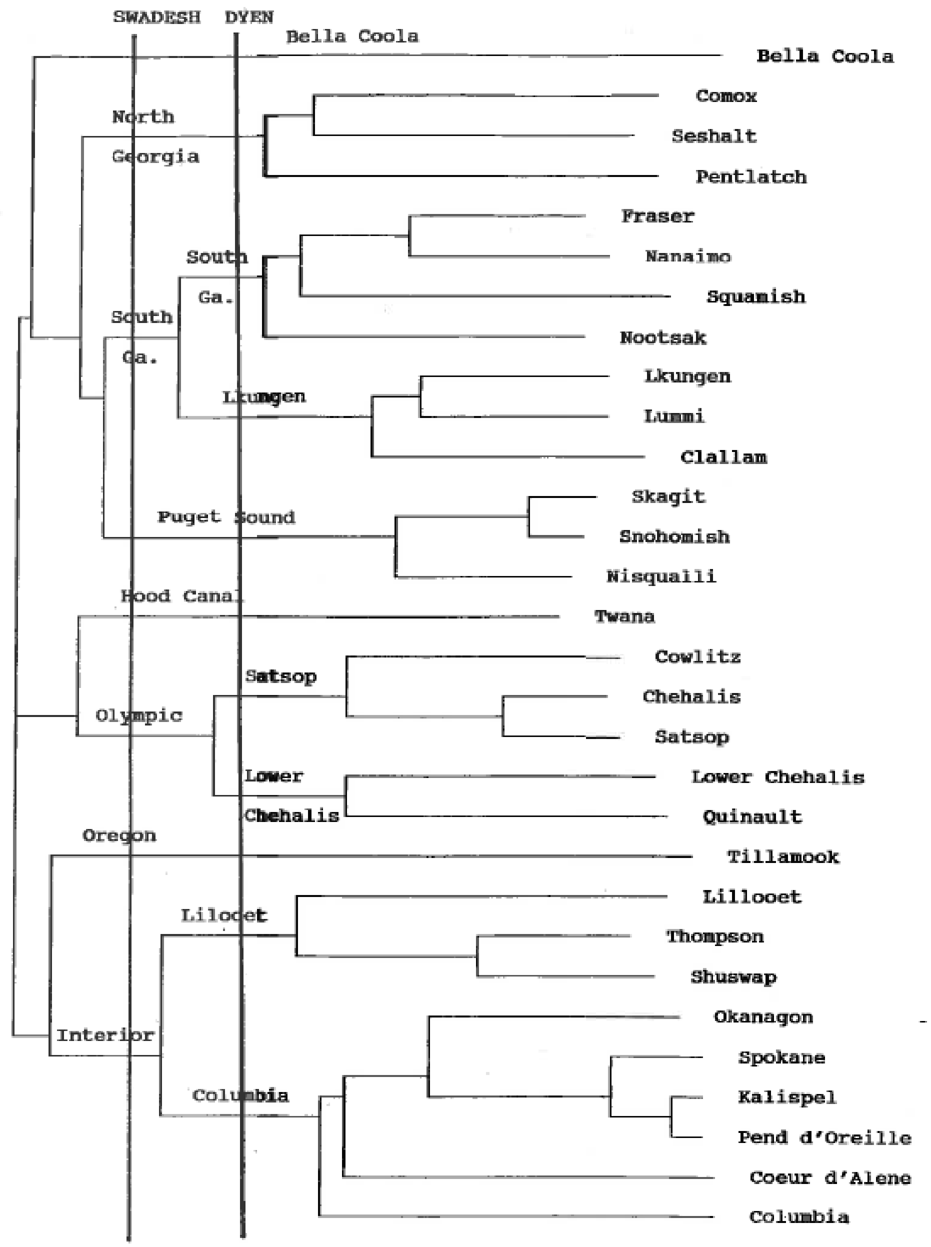
(ref. Shepard 1980)

Additive tree of generalization data on societal risks (Johnson & Tversky, 1987)



Fitting an additive tree to derive useful partitions and compare them (Corter 1996)

Data are linguistic similarities of Native American languages, Pacific Northwest area (taxonomies of Swadesh and Dyen shown as vertical lines).



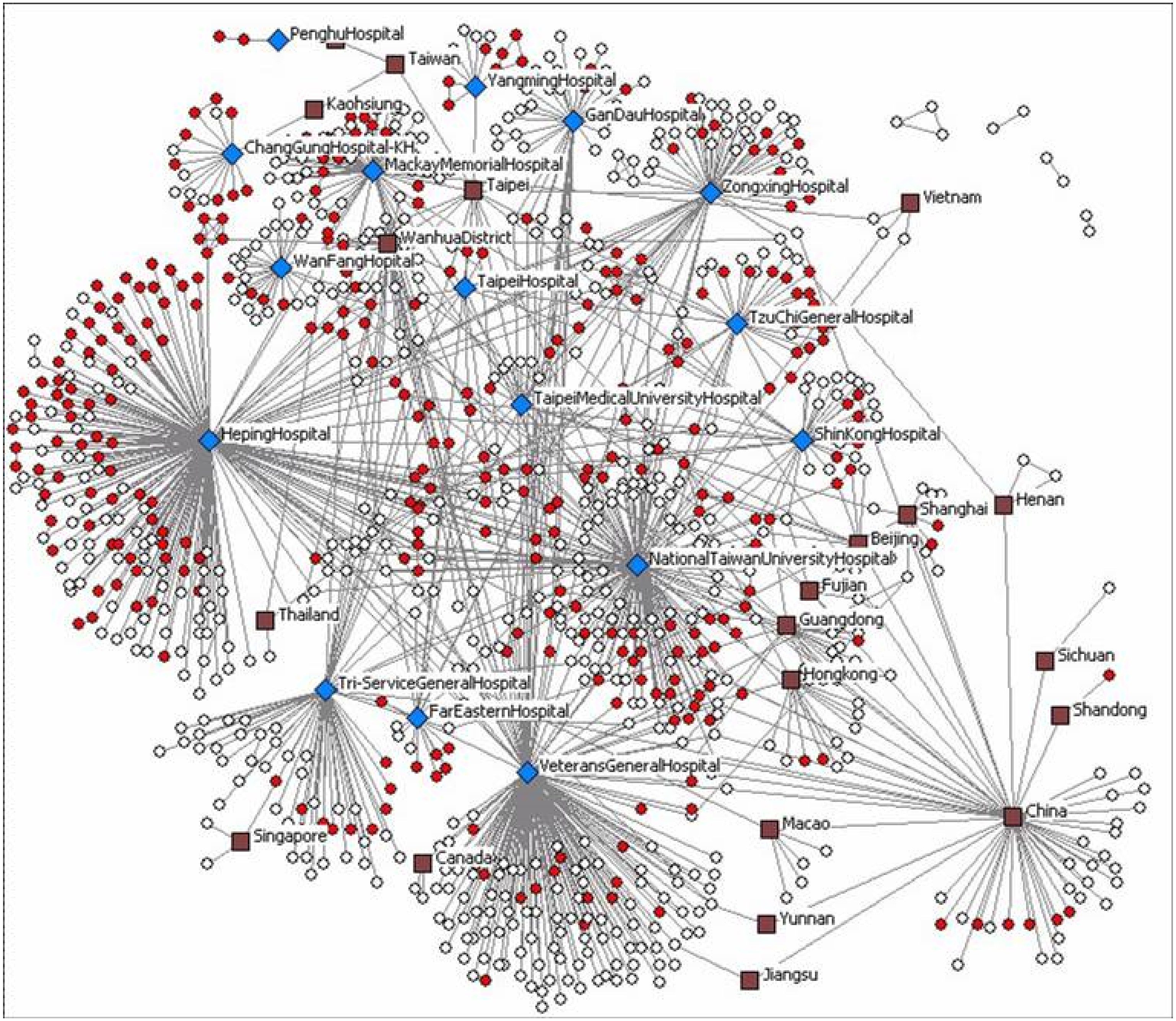
Social nominations data: finding an optimal organizational hierarchy (from Corter, 1996)

TABLE 3: Summed number of nominations of column departments by respondents in row department.

dept:	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	n
01	0	0	1	2	1	1	0	0	2	1	1	1	0	2	0	1	0	2
02	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	2
03	0	0	0	2	1	1	0	0	0	1	1	0	0	1	0	0	0	2
04	3	0	0	0	1	2	0	1	3	3	0	2	0	1	0	3	0	3
05	0	3	3	2	1	0	0	0	1	1	3	1	0	0	4	3	3	5
06	0	0	1	1	0	0	0	4	0	0	1	0	0	3	0	1	0	4
07	0	0	0	0	0	0	0	0	0	1	0	2	1	0	0	0	1	2
08	0	0	0	1	1	1	0	0	0	0	0	0	0	1	1	0	0	2
09	2	0	1	3	1	0	0	2	1	0	0	0	0	2	0	1	1	6
10	2	1	3	1	1	0	2	1	2	0	2	2	1	1	1	1	1	3
11	0	3	0	1	3	1	0	0	1	2	0	0	1	0	3	0	0	4
12	1	0	0	1	0	0	3	0	1	3	0	0	1	0	0	1	1	3
13	0	2	0	1	0	2	2	1	0	0	2	1	0	0	2	1	1	2
14	0	0	0	1	0	1	1	1	1	0	1	0	1	0	1	0	0	4
15	0	6	0	0	6	0	0	1	0	0	4	0	0	0	0	0	0	7
16	0	0	0	3	3	2	0	0	1	1	0	0	0	0	0	0	2	5
17	0	1	0	0	1	0	1	0	0	0	1	1	0	0	1	2	0	2

num	label	div	full name of department
01	arts_edu	IV	The Arts in Education
02	clin_psy	II	Clinical Psychology
03	comput_e	IV	Communication, Computing, and Technology in Education
04	curric_t	III	Curriculum and Teaching
05	devel_ed	II	Developmental and Educational Psychology
06	ed_admin	III	Educational Administration
07	hlth_nut	V	Health and Nutrition Education
08	adult_ed	III	Higher and Adult Education
09	lang_lit	IV	Languages, Literature, and Social Studies in Education
10	math_sci	IV	Mathematics and Science Education
11	measrmnt	II	Measurement, Evaluation, and Applied Statistics
12	move_sci	IV	Movement Sciences and Education
13	nurse_ed	V	Nursing Education
14	phil_soc	I	Philosophy and the Social Sciences
15	soc_cnsl	II	Social, Organizational, and Counseling Psychology
16	specl_ed	III	Special Education
17	speech_p	II	Speech and Language Pathology and Audiology

<u>Social</u>
<u>Networks</u>:
Graph
models of
association
data
among
hospitals



From
Arizona
AI lab:

Analysis of social interactions at a conference:

.. 70 participants (60 from the sponsor companies, 10 from the MIT Media Lab) wore Sociometric badges during the event.. badge then sent the information over USB to a database, which was read out by a social network visualization program (a modification of the GUESS system developed by Eytan Adar)

..

[Right: SN Diagram after the last break]

From almost the very beginning there was one giant component with a strong core-periphery structure.. It appears that there were two factors that led to this structure:

1. Media Lab participants, who all spoke to each other and spoke with many sponsor companies
2. Research affiliates: members of sponsor companies who had also worked at the Media Lab as visiting researchers. These participants knew other research affiliates who had been at the Media Lab from different companies at the same time as well as the Media Lab participants.