

## Clusteranalyse

*Hartmut-A. Oldenbürger*

### *1. Zur Entwicklung der Literatur*

(Aufgabenstellung der vorliegenden Arbeit)

Clusteranalytische Verfahren konstruieren Gruppierungen von Objekten aufgrund ihrer ‚Ähnlichkeit‘.

Obwohl Methoden zur Realisierung dieses Ziels schon um 1935 (Tryon, 1939) vorgeschlagen wurden, fanden sie (im Vergleich zur weit verbreiteten Anwendung der Faktorenanalyse) wenig Beachtung. Erst die Veröffentlichung der inzwischen klassischen Monographie von Sokal und Sneath (1963) und gleichzeitig die größere Verfügbarkeit leistungsfähiger Rechenanlagen gaben die Anstöße zur breiteren Methodenentwicklung und expandierendem Einsatz clusteranalytischer Verfahren zunächst in der systematischen Biologie (Taxonomie), später auch in anderen empirischen Wissenschaften.

Für die Psychologie hat der Aufsatz ‚Hierarchical Clustering Schemes‘ von Stephen C. Johnson (1967) den entscheidenden Impuls zur methodologischen Diskussion und Anwendung von Clusteranalysen gegeben (Blashfield, 1980). Sie treten nun häufig an die Stelle von Faktorenanalysen, die in der Forschungspraxis sehr oft - inkorrekterweise - zur Generierung von Variablen-**gruppierungen** eingesetzt worden sind. (Die Angemessenheit dieser Vorgehensweise insbesondere für die Skalenkonstruktion mit Hilfe von Tests diskutiert Revelle (1978, 1979); ein vergleichendes Beispiel für die WISC-R gibt Silverstein, 1980.)

Inzwischen kann die Entwicklung des Umfangs der Literatur als explosionsartig bezeichnet werden: Während in anderen Bereichen die Verdoppelungszeit der Publikationszahlen bei zwölf bis fünfzehn Jahren angesetzt werden kann, liegt sie für die clusteranalytische Literatur z.Zt. bei drei bis vier Jahren (Blashfield/Aldenderfer, 1978). Dieser Sachverhalt zeitigt unerfreuliche Begleiterscheinungen: (a) es bilden sich Cliques von Anwendern der Cluster-

analyse, die den Einsatz nur bestimmter Verfahren tradieren und (b) die Ausbildung gegenstandsbereichsspezifischer Jargons behindert die interdisziplinäre Kommunikation (Soweit die Ergebnisse einer Benutzerbefragung und einer Analyse des ‚Science Citation Index‘ von Blashfield und Aldenderfer, 1978.).

Davon scheint zumindest nach 1975 auch die methodologisch orientierte clusteranalytische Literatur betroffen zu sein. Während die wichtigen Lehrbücher und Monographien von Anderberg (1973), Bock (1974), Duran/Odell (1974), Hartigan (1975), Jardine/Sibson (1971), Sneath/Sokal (1973) und Vogel (1975) zumindest in ihren bibliographischen Teilen die verfügbare Literatur umfassend aufzuarbeiten versuchten, erscheinen nach 1975 gehäuft in die Verfahren einführende Artikel (z.B. Kopp 1978a, b, c) und elementare Lehrbücher, die die Fortentwicklungen der clusteranalytischen Methoden und Überprüfungsansätze nahezu unberücksichtigt lassen (vgl. dazu Steinhausen/Langer, 1977, Opitz 1980). 1980 allerdings wurde der beachtenswerte Studententext ‚Clusteranalysen‘ von Thomas Eckes und Helmut Roßbach publiziert. Dieser schließt die bestehende Lücke, indem er die wichtigsten methodologischen Ansätze des Gebiets aufarbeitet. Er mußte deshalb auf Intention, Planung und Gestaltung dieser Arbeit Einfluß haben. Der vorliegende Artikel erhebt danach keinerlei monographischen Anspruch, er soll stattdessen

- (a) eine gestraffte Übersicht der wesentlichsten clusteranalytischen Problemstellungen und Verfahren (mit den zugehörigen Referenzen), sowie
- (b) Ergänzungen und Hinweise zu methodologischen Weiterentwicklungen von Ansätzen und grundlegenden Innovationen geben.

Auf eine ausführliche Darstellung der mathematischen Grundlagen und der Vorgehensweise (an Beispielen und Veranschaulichungen) wird im allgemeinen verzichtet werden. (Siehe dazu Eckes/Roßbach, 1980. - Auch die kombinatorischen Ausdrücke zur Angabe der Anzahlen möglicher Lösungen von clusteranalytischen Problemstellungen werden nicht mitgeteilt. Diese führen zumeist lediglich zu der Aussage, das gestellte Problem sei wegen der immensen Anzahl der Möglichkeiten nicht exakt lösbar (siehe dazu Duran/Odell, 1974 und verschiedene Passagen in Bock, 1974 und Flachsmeier, 1972); die Existenz von ‚backtracking-‘ und ‚branch and bound-‘ Verfahren aber belegt die Vordergründigkeit dieses Arguments.)

Einerseits soll also dem unbelasteten Leser zum Einstieg eine Strukturierung des Bereichs gegeben werden (advanced organizer), andererseits soll der kenntnisreichere Leser Hinweise auf und Anstöße zu Weiterentwicklungen erhalten.

## 2. Zur Datenerhebung und Datenstruktur

Dieses Kapitel setzt sich nicht mit genuin clusteranalytischen Fragestellungen auseinander. Vor der Darstellung der Verfahrensklassen (in Kapitel 3) erfolgt zunächst eine ordnende Charakterisierung des Datenmaterials und seiner Erhebung. Dazu werden einige notwendige Begriffe eingeführt und mit Hinweisen auf für die Psychologie kritische (Interpretations-)Probleme verknüpft.

Technisch gesehen liegen am Ausgangspunkt einer clusteranalytischen Fragestellung Daten gewöhnlich in (mindestens) einem der folgenden Modi vor:

- (a) als  $n \times m$  - Matrix  $[x_{ij}]$ ; durch die Eintragungen/Zellen dieser Matrix wird jeder Beobachtungswiederholung  $i$  eine Ausprägung eines Merkmals  $j$  (numerisch oder nichtnumerisch) zugeordnet. (Beispiele für Beobachtungswiederholungen sind Gegenstände, Personen, Situationen etc.; Beispiele für Merkmale sind Attribute, nominale Klassen, Variablen etc.) (Speziell im Fall numerisch interpretierter Ausprägungen der Merkmale können die Beobachtungswiederholungen statistisch als Realisationen von Zufallsvariablen angesehen oder als Punkte in einem (nicht notwendig euklidischen) mehrdimensionalen Raum aufgefaßt werden. Diese Sichtweisen sind auch für die andere Modalität, nämlich der Merkmale, einsetzbar.) Die Matrix  $[x_{ij}]$  wird hier häufig einfach als ‚Datenmatrix‘ bezeichnet.
- (b) als  $r \times r$  - Matrix  $[p_{kl}]$ ; durch die Eintragungen dieser Matrix wird jedem Objektpaar  $(o_k, o_l)$  aus einer Objektmenge  $O = \{o_1, \dots, o_k, o_l, \dots, o_r\}$  eine reelle Zahl  $p(o_k, o_l)$  oder kurz  $p_{kl}$ , zugeordnet, die die ‚Proximity‘ der beiden Objekte angibt. Der allgemeine Begriff der ‚Proximity‘ steht für mehrere Klassen von Maßen, etwa für die ‚Ähnlichkeit‘ bzw. ‚Unähnlichkeit‘, die ‚Nähe‘ bzw. ‚Entfernung (Distanz)‘, den ‚Zusammenhang‘ oder die ‚Abhängigkeit‘, die ‚Kovariation‘ oder ‚Interaktion‘ von Objektpaaren. (Beispiele für Objekte sind (Zufalls-)Variablen, Gegenstände, Lebewesen, Personen, Ereignisse, Situationen, Gegenstände der Anschauung oder des Denkens (Perzepte, Konzepte) etc. - Beispiele für Proximity-Maße sind Distanzen, Kovarianzen, Korrelationen, Maße erklärter Varianz, Kontingenzkoeffizienten, Häufigkeiten gemeinsamen Auftretens (insbesondere bei Ereignissen), Produktwahrscheinlichkeiten, bedingte Wahrscheinlichkeiten, Verwechslungswahrscheinlichkeiten, subjektive Ähnlichkeitsurteile etc.) Die Matrix  $[p_{kl}]$  wird hier häufig einfach als ‚Proximitymatrix‘ bezeichnet.

Da clusteranalytische Verfahren Gruppierungen zusammengehöriger Objekte zumeist auf der Basis von Proximitymatrizen generieren, ist es für die theoretisch-inhaltliche Interpretation der Resultate von herausragender Bedeutung, auf welchem Wege die Proximitymaße gewonnen wurden.

- (a) direkt: Im Rahmen z.B. einer Konkretisationserhebung kognitiver Strukturen werden Probanden nach ihrem subjektiven Urteil zur (paarweisen) Ähnlichkeit oder Zusammengehörigkeit von Perzepten oder Konzepten befragt. (Übliche Erhebungsparadigmata: Ratio-Scaling, (vollständiger) Paarvergleich, Sortierverfahren etc. - In diesem Fall ist es unzureichend, das Ergebnis der Erhebung, die Proximities, lediglich durch Repräsentationsverfahren (MDS und/oder Clusteranalyse) zu beschreiben und zu interpretieren, vielmehr ist ebenso eine theoretische Bemühung um die Explikation des Urteilsprozesses als kritisches Problem anzusehen.) Theoretische Aspekte der Ähnlichkeitsmessung erörtern Eckes/Roßbach, 1980, S. 36f..
- (b) indirekt: Im Rahmen z.B. einer Fragebogenuntersuchung werden Probanden aufgefordert, zu verschiedenen Statements die jeweilige Ausprägung ihrer persönlichen Zustimmung bzw. Ablehnung zu skalieren. Ergebnis dieses Vorgehens ist eine  $n \times m$  - Datenmatrix  $[x_{ij}]$  (Probanden  $\times$  Statements). Will man eine kompakte Beschreibung der Datenstruktur z.B. durch Gruppierung herstellen, so ist zunächst - auf dem Begründungshintergrund der Fragestellung - die Modalität zu wählen, deren Elemente (Personen bzw. Items) als Objekte einer Clusteranalyse dienen sollen. Ferner ist begründet ein geeignetes Maß zu wählen oder zu konstruieren, welches die Proximity der Objekte beschreiben soll. Dieses Maß ist für das Resultat des Repräsentationsverfahrens (z.B. Clusteranalyse) konstitutiv. - Für die theoretisch-inhaltliche Einordnung des Gesamtergebnisses ist hervorzuheben, daß bei diesem Vorgehen die Ausprägung der ‚Ähnlichkeiten‘ bzw. ‚Unähnlichkeiten‘ zwischen den Objektpaaren nicht direkt erhoben, sondern errechnet wurde. Dabei benutzt man implizit bestimmte Variationsquellen (z.B. für die Korrelation zwischen Items die interindividuelle Varianz). Dieser Sachverhalt ist bei der Interpretation der Resultate als kritisches Problem anzusehen. (So wäre es für das obige Beispiel fehlerhaft, die durch Verfahrenseinsatz konstruierte Gruppierung der Items als Urteilsstruktur eines mittleren, ‚idealen‘ Individuums auszuzeichnen.)

Das Problem der Überführung der in der Forschungspraxis häufig vorliegenden Datenmatrix  $[x_{ij}]$  in eine Proximitymatrix  $[p_{kl}]$  kann nicht durch generelle Empfehlung gelöst werden. Auch innerhalb der Klassen der für numerisch interpretierbare, nominale und gemischte Daten angemessenen bivariaten Zusammenhangs- und Entfernungsmaße, werden durch die vielfältigen Koeffizienten verschiedene Aspekte der Kovariation formalisiert bzw. gemessen. Es ist deshalb im jeweiligen empirischen Kontext zu explizieren, unter welcher Fragestellung bzw. welchem Ziel das spezifische Proximitymaß vor dem Einsatz eines Repräsentationsverfahrens gewählt wurde. Die häufig geübte Praxis der voraussetzungsunkritischen, automatisierten Verwendung von Standardmaßen (z.B. Korrelationskoeffizient) ohne exploratorische Bemühungen (z.B. zur Kontrolle möglicher Artefaktquellen) ist unangemessen.

Darstellungen der verschiedenen Zusammenhangs- und Entfernungsmaße finden sich u.a. bei Anderberg (1973, S. 72ff.), Bock (1974, S. 24ff.), Sneath/Sokal (1973, S. 114ff.), Steinhausen/Langer (1977, S. 51 ff.); die wohl ausführlichste Übersicht zu symmetrischen und unsymmetrischen Kovariationsmaßen speziell für binäre Variablen findet sich bei Kleiter/Petermann (1977, S. 43ff.).

Als spezielle Proximitymaße sind **Distanzfunktionen** auch für clusteranalytische Verfahren von besonderer Bedeutung. (Sie dienen oft als Kriterium zur Messung der Homogenität von Objekten innerhalb von Gruppierungen bzw. zur Messung der Heterogenität von Objekten, die verschiedenen Gruppierungen angehören.) Deshalb werden die Eigenschaften dieser Maße hier im Rahmen einer Definition aufgeführt:

Seien  $o_k, o_l, o_s$  Elemente von  $O$ . Eine Funktion  $d: O \times O \rightarrow \mathbb{R}$  heißt Distanzmaß, genau dann wenn für alle  $o_k, o_l, o_s \in O$  gilt:

- |       |  |                     |
|-------|--|---------------------|
| (D.1) | $d(o_k, o_l) \geq 0$                         | Positivität         |
| (D.2) | $d(o_k, o_l) = 0 \Leftrightarrow o_k = o_l$  | Identität           |
| (D.3) | $d(o_k, o_l) = d(o_l, o_k)$                  | Symmetrie           |
| (D.4) | $d(o_k, o_s) \leq d(o_k, o_l) + d(o_l, o_s)$ | Dreiecksungleichung |

Das Paar  $(O, d)$  heißt metrischer Raum;  $d$  wird auch Metrik genannt.

Gilt für alle  $o_k, o_l, o_s \in O$  die folgende Verschärfung von (D.4):

$$(D.5) \quad d(o_k, o_s) \leq \max \{d(o_k, o_l), d(o_l, o_s)\}$$

so heißt  $d$  **Ultrametrik**. (Anschaulich besagt (D.5), daß alle Tripel von Objekten gleichschenklige Dreiecke aufspannen, deren Basis kleiner ist als die gleichlangen Schenkel.) Johnson (1967) hat auf die besondere Bedeutung der Ultrametrik für einige Methoden der hierarchischen Clusteranalyse (siehe 3.3) hingewiesen: Diese können als Transformationsverfahren von empirischen **Unähnlichkeitsmaßen** (Geltung von mindestens (D.1) bis (D.3)) in Ultrametrien angesehen werden.

Da ultrametrische Distanzmaße bei Vorliegen einer Datenmatrix  $[x_{ij}]$  i. a. nicht direkt berechnet werden können, aber eine Reihe von Proximitymaßen die Distanzaxiome (D.1) bis (D.4) erfüllen oder metrische Äquivalente besitzen, erscheint es sinnvoll - soweit es von der Fragestellung her angemessen ist - zur Transformation von Datenmatrizen  $[x_{ij}]$  in Proximitymatrizen  $[p_{kl}]$  möglichst Metriken zu verwenden.

Bei Vorliegen numerisch interpretierbarer Eintragungen der Datenmatrix stellt die folgende Funktion eines der prominentesten Beispiele für eine Metrik dar:

$$d(o_k, o_l) = \left( \sum_j |x_{kj} - x_{lj}|^r \right)^{1/r}; \quad r \geq 1$$

Diese Distanzfunktion heißt allgemein  $L_r$ -Metrik (in der Literatur zur mehrdimensionalen Skalierung: Minkowski-r-Metrik). Wird der Exponent  $r$  spezifiziert, so entstehen als ausgezeichnete Fälle:

- für  $r = 1$ , die City-Block-Metrik
- für  $r = 2$ , die euklidische Metrik
- für  $r = m$ , die Supremum-Metrik

(Zu weiteren Eigenschaften dieser Abstandsfunktionen siehe Eckes/Roßbach 1980, S. 42ff. - Im originären Zusammenhang mit der mehrdimensionalen Skalierung stehen die Erörterungen von Ahrens 1974, z.B. S. 190ff.) Weitere Entfernungsmaße, die z.B. auch bei Vorliegen binärer oder nominaler Merkmalsausprägungen eingesetzt werden können, finden sich in der zu den Proximitätsmaßen genannten Literatur.

Damit sind die wichtigsten Begriffe und Überlegungen zur Charakterisierung der Ausgangslage beim Einsatz clusteranalytischer Verfahren eingeführt. Im nächsten Kapitel wird das Vorliegen von Daten- und/oder Proximitymatrizen vorausgesetzt.

### 3. Problemstellungen und Verfahren

#### 3.1 Untermengenauswahl

Bemerkenswerterweise tritt das Problem der Selektion von Elementen aus einer gegebenen Menge unter Optimierung eines Kriteriums in verschiedenen methodischen und inhaltlichen Bereichen häufig auf, es wurde aber m. W. in der clusteranalytischen Literatur bisher nicht als eigenständige Fragestellung identifiziert. Deshalb seien einige Beispiele und Hinweise gegeben, die die Problemstellung charakterisieren und umreißen, sowie mögliche Lösungen andeuten:

- (a) Aus einer Menge von Prädiktoren mit dem Umfang  $n$  soll eine  $k$ -elementige (echte) Untermenge selegiert werden, so daß eine Kriteriumsvariable maximal prädiziert werden kann (Optimierungskriterium: z. B. adjustiertes  $R^2$ ). Dies ist das Standardproblem der multiplen Regressionsanalyse.
- (b) Aus einer Menge von Variablen soll eine Untermenge vom Umfang  $k$  ausgewählt werden, so daß die Varianz aller restlichen Variablen maximal erklärt werden kann. Dieses Verfahren identifiziert die Stellvertreter eines Variablenkollektivs; es wird Interdependenzanalyse genannt (Boyce/Farhi/Weisedel, 1974, S. 67ff.) und kann als attraktive Alternative zur Faktorenanalyse angesehen werden. (Zwar entfällt hier das Problem der Faktorinterpretation, es wird aber ein gewisser Verlust an Varianzerklärung in Kauf genommen.)

- (c) Aus einer Menge von Gegenständen mit gegebenen Eigenschaften sollen Repräsentanten für die verbleibenden Objekte ausgewählt werden. Während die Interdependenzanalyse Merkmale (Variable) selektiert, soll nun nach zentralen Merkmalsträgern (Wechsel der Modalität) gesucht werden. Handelt es sich bei den Gegenständen z.B. um Perzepte oder Konzepte, so ist dies die Frage nach den Prototypen (z.B. von Kategorien), auf deren ausgezeichnete Rolle insbesondere Rasch (1975, 1978) hingewiesen hat.
- (d) Aus einer Menge von Gegenständen mit (numerischen) Merkmalen soll eine Untermenge selektiert werden, deren Homogenität gegenüber dem Rest besonders groß ist. Diese Separation von Clustern vom Hintergrund (Yahil/Brown, 1976) kann in verschiedensten Anwendungsbereichen in Frage stehen, z.B. in der Astronomie bei der Herauslösung von Galaxien aus diffusen Verteilungen von Sternen, in der Medizin bei der Diagnostik von Auffälligkeiten in Röntgenbildern (Block et al., 1974) oder z.B. in der Psychologie bei der Identifikation besonders merkmals homogener Personen in einer Gesamtgruppe. (Weitere Stichworte: Mustererkennung, Kategorienbildung, Klassifikationsprozesse.)
- (e) In engstem Zusammenhang mit der im vorigen Punkt angesprochenen Problemlage steht die Fragestellung der Analyse von Mischverteilungen. (Dieser Ansatz könnte auch unter Abschnitt 3.2 eingeordnet werden.) Man geht dabei von der Vorstellung aus, ein vorliegendes Sample sei aus einer Population gezogen, die aus mindestens zwei Subpopulationen mit Zufallsvariablen unterschiedlicher Mittelwertsvektoren und/oder Kovarianzmatrizen besteht. Aufgabe der Analyse ist die Bestimmung der Anzahl von Subpopulationen, die Schätzung der Parameter der multivariaten Zufallsvariablen und die Zuordnung der Beobachtungswiederholungen (Realisationen) zu den Subpopulationen. (Siehe Bock, 1974, S. 250ff.; Hartigan, 1975, S. 113ff.; Eckes/Roßbach, 1980, S. 92f.; Wolfe, 1970, 1978.)  
Der abschreckenden Komplexität dieser Aufgabenstellung steht die Attraktivität einer möglichen Lösung gegenüber: für nahezu alle Verfahren der multivariaten Datenanalyse (Regressions-, Varianz-, Interdependenz-, Faktor-, Diskriminanzanalyse etc.) stünde ein Überprüfungsmedium der Homogenität der Stichprobe zur Verfügung.
- (f) Aus einer Menge von Objekten mit gegebener Proximitymatrix soll eine Untermenge selektiert werden, deren Proximities einer spezifischen - algebraisch formulierten - Charakteristik, z.B. einem System von (Un-) Gleichungen, möglichst gut genügen. Beispiele für derartige Problemstellungen sind etwa: die Auswahl Guttman- bzw. Rasch-skalierbarer Items aus einem (vorläufigen) Test; die Auswahl von Variablen mit paarweise auffällig linearem Zusammenhang; die Auswahl von Gegenständen der Anschauung oder des Denkens mit symmetrischer Ähnlichkeit oder die

Auswahl von Objekten, deren Proximities durch eine Ultrametrik besonders gut dargestellt werden können (Hubert, 1980). (Da sich solche Vorgehensweisen einem Immunisierungsverdacht besonders stark aussetzen, ist für sie die Entwicklung und Durchführung angemessener Überprüfungsverfahren dringend angezeigt.)

Diese kurze Übersicht soll deutlich machen: Gemeinsam ist den Ansätzen die Aufgabenstellung der Untermengenselektion, sie unterscheiden sich aber gravierend in ihren Zielsetzungen, die durch mathematische Formulierung von Optimalitätskriterien, sog. Zielfunktionen, zu explizieren sind (Beale, 1970). Die Verschiedenartigkeit der Zielsetzungen, die sich in der Heterogenität der Optimalitätskriterien hinsichtlich ihrer mathematischen Eigenschaften widerspiegelt, läßt es nicht erwarten, daß zur Lösung des Auswahlproblems eine allgemein anwendbare optimale Verfahrensvorschrift (Algorithmus) angebar ist (abgesehen von der Möglichkeit der Totalenumeration für uninteressant kleine Unter- bzw. Obermengen).

Für verschiedene Problemstellungen allerdings liegen in der methodologischen Literatur, von der Forschungspraxis m. W. bisher unbeachtet, effiziente Lösungsverfahren vor: Schon 1967 publizierten Beale, Kendall und Mann einen Algorithmus, der für die unter den Punkten (a) bis (c) angesprochenen Aufgaben optimale Lösungen liefert. (Demgegenüber sind z.B. die von den verbreiteten Programmpaketen BMDP und SPSS realisierten Vorgehensweisen und Resultate (!) bei der Variablenauswahl der multiplen Regression als suboptimal zu bezeichnen.) Die programmtechnische Umsetzung des Verfahrens leisteten Boyce, Farhi und Weisedel (1974). (Sie weisen z.B. auch die Überlegenheit des Beale/Kendall/Mann-Algorithmus gegenüber den Standardauswahlverfahren der multiplen Regression (forward inclusion, backward elimination) aus. - In das Programmpaket BMDP-77 (siehe Dixon/Brown, 1977, S. 418ff.) wurde ein Algorithmus von Furnival/Wilson (1974) eingesetzt; siehe auch McHenry (1978).)

Für die komplexe Aufgabenstellung der Analyse von Mischverteilungen sei ergänzend auf das Vorliegen einer (weiteren) programmsprachlichen Formulierung (Fortran-Code) eines Verfahrens bei Hartigan (1975, S. 125ff.) hingewiesen. Hubert (1980) diskutiert eine Strategie der Datenanalyse zur Identifikation von Untermengen, für deren Elemente die Proximities bestimmten algebraischen Eigenschaften möglichst gut gehorchen.

Die methodischen Vorgehensweisen und Lösungsansätze zur optimalen Untermengenauswahl wurden bisher kaum beachtet; sie warten auf Bewährungsprüfungen.

### 3.2 Mengenzerlegung

Als Partition bzw. Zerlegung einer Menge  $0$  bezeichnet man eine Menge (Familie)  $Z(0)$ , deren Elemente (nichtleere) Untermengen von  $0$  sind:  $G_p, G_q, \dots$  und folgende Forderungen erfüllen:

$$(Z.1) \quad p \neq q \Rightarrow G_p \cap G_q = \emptyset$$

$$(Z.2) \quad G_1 \cup G_2 \cup \dots \cup G_p \cup \dots = \bigcup_p G_p = 0,$$

d.h. die in  $Z(0)$  als Elemente enthaltenen Mengen  $G_p, G_q$ , die hier Gruppierungen genannt werden, sind paarweise elementenfremd bzw. disjunkt und sie erschöpfen die Menge  $0$  bzw. decken diese vollständig ab.

Zwei (triviale) Zerlegungen einer Menge  $0$  sind als extreme Fälle besonders ausgezeichnet: die feinste Partition (engl. ‚Splitter‘) enthält als Elemente alle einelementigen Untermengen von  $0$ ; die größte Partition (engl. ‚lumber‘) enthält als Element nur die Menge  $0$  selbst.

Als Teilklasse clusteranalytischer Verfahren hat die Partitionierung generell folgende Aufgabenstellung: Man konstruiere bzw. suche zu einer gegebenen Objektmenge eine Zerlegung mit Gruppierungen  $G_p$ , für deren Elemente gilt: Gehören zwei Objekte zu gleichen Gruppierungen, so sollen sie einander möglichst ähnlich sein (innerhalb - homogen), gehören sie zu verschiedenen Gruppierungen, so sollen sie untereinander möglichst unähnlich sein (zwischen - heterogen). Als Ergebnis der Verfahrensanwendung wird die Feststellung der Anzahl der Gruppierungen und die Zuordnung der Objekte zu den Gruppen erwartet.

Die große Vielfalt der Methoden der disjunkten Gruppierung unterscheidet sich hauptsächlich danach,

- auf welche Informationen (Datenmatrix mit numerisch oder nicht numerisch interpretierbaren Merkmalsausprägungen und/oder Proximitymatrix) zurückgegriffen wird bzw. werden kann,
- ob und wie ein Optimierungskriterium (Zielfunktion) mathematisch spezifiziert wurde und
- nach welchem verfahrenstechnischen Konstruktionsprinzip die Methoden während des Analyseprozesses die Zerlegungen generieren (Enumeration; Modi der Zuordnung von Einzelobjekten zu bestehenden Gruppen, des Austausches von Objekten zwischen den Gruppen sowie der Separation und/oder Fusion von Gruppen).

Von verbreitetem forschungspraktischen Interesse sind Anwendungen, in denen zu den Objekten nur ein numerisches **Merkmal** gegeben ist oder betrachtet

wird (univariate Klassifikation, z.B. Aufteilung einer Personengruppe nach Intelligenz). In solchen Situationen ist zwar eine Aufteilung nach Perzentilen (Median, Terzile, Quartile etc.) gängig, diese läßt aber die Verteilungsform der Variablen unberücksichtigt. So geht beispielsweise die Trennung am Median für nicht-symmetrische Verteilungen mit einem (vermeidbaren) Informationsverlust einher. Eine optimale Aufteilung nach festgelegter Gruppenzahl und nahezu beliebig wählbarem Optimierungskriterium (z.B. Varianz zwischen den Gruppen) ist aber praktisch immer exakt möglich. Dies leistet ein Algorithmus von W. D. Fisher (1958), dessen Einsatz z.B. zur Variablenkategorisierung vor Durchführung einer Konfigurationsfrequenzanalyse angezeigt ist. Eine programmsprachliche Formulierung (Fortran-Code) des Verfahrens teilen Anderberg (1973, S. 228ff.) und Hartigan (1975, S. 141) mit.

Exakte Partitionierung (z.B. ebenfalls mit dem Fisher-Algorithmus) ist auch möglich, wenn die Aufteilung zwar nach mehreren Variablen (simultan) vorgenommen werden soll, die Anzahl der möglichen Partitionen aber dadurch erheblich eingeschränkt ist, daß den Objekten hinsichtlich einer zusätzlichen Variablen eine natürliche Ordnung unterliegt oder unterstellt wird (z.B. im Fall einer Konstruktion von Stufen der Entwicklung hinsichtlich mehrerer kognitiver Merkmale, wobei die **„natürliche Reihenfolge“** nach der Variablen ‚Alter‘ eingehalten werden soll).

Im generellen Anwendungsfall, bei Vorliegen **multivariater Daten und/oder von Proximitymatrizen**, ist die Angabe von Verfahren, die zu exakten Partitionierungen führen, also eine Optimierung des spezifizierten Kriteriums garantieren (Rao, 1971), nicht möglich. (Abgesehen von den seltenen Fällen, in denen wegen des geringen Umfangs der Objektmenge und/oder einer Festlegung auf sehr wenige Gruppierungen eine Totalenumeration der Partitionen durchführbar und sinnvoll ist, siehe Duran/Odell, 1974, S. 32ff.) Deshalb werden heuristische Methoden verwendet, die z.B. von Startpartitionen ausgehen und diese durch Austausch von Elementen zwischen den Gruppen und Verschmelzen bzw. Aufteilen von Gruppen hinsichtlich des Wertes einer Zielfunktion sukzessive verbessern. Systematische Darstellungen zu Konstruktionsmethoden von (lage-)homogenen Gruppierungen, die das Varianzkriterium, die Kriterien der Spur oder Determinante der Kovarianzmatrix oder das Kriterium der Abstandsquadratsumme (Büttner, 1975) optimieren, werden von Bock (1974, S. 104ff.), Hartigan (1975, S. 74ff.), Späth (1975, S. 34ff.) und Steinhausen/Langer (1977, S. 100ff.) gegeben. Eine theoretisch orientierte Übersicht im Rahmen der Theorie(n) der kombinatorischen Programmierung liefern Balas/Padberg 1975.

Ergänzend sei auf die folgenden Weiterentwicklungen verwiesen:

- Ein effizienter (branch-and-bound-) Algorithmus zur globalen Optimierung des Varianzkriteriums wird von Herden/Steinhausen (1979, mit Fortran-Programm) mitgeteilt.

- Da viele Methoden den Nachteil haben, zur Konstruktion von Zerlegungen mit relativ gleich großen Gruppen zu tendieren, ist ein Verfahrensvorschlag von Eyes (1977a, 1977b, 1978) von besonderem Interesse: Grundidee ist dabei die Umschreibung lokaler Dichtezentren mit Hyperkörpern (Quadern, Ellipsoiden), die in ihrer Größe variieren und deren Lage im Raum verschoben wird. - Der oben angesprochene Nachteil kann auch mit Hilfe eines skalierungsinvarianten Partitionsverfahrens von Späth/Müller (1979) vermieden werden.
- Da die meisten Verfahren auf euklidische Optimierungskriterien zurückgreifen, sind sie anfällig gegen die Verletzung der normalerweise nicht zutreffenden Voraussetzung der Normalverteilung und gegen Vorliegen von Ausreißern (Outlier). Dies kann durch den Einsatz von resistenteren bzw. robusteren Methoden vermieden werden, die z.B.  $L_1$ -Kriterien optimieren (siehe Späth, 1976, mit Fortran-Programm).
- Überlegungen zu Partitionierungsverfahren bei Vorliegen großer Objektmengen ( $n > 100$ ) stellen Marsten (1975) und Späth (1977a mit Fortran-Programm) an.

Zur **Evaluation von Partitionierungsverfahren** (Rand, 1971) legte Milligan (1980) eine empirische Untersuchung von vier K-Means-Algorithmen vor. Es wurden Datensätze mit bekannter Gruppierung durch sechs Arten von Fehlern überlagert und die Reklassifizierungsleistung der Verfahren erhoben. Diese erwies sich als sehr gut, wenn den Verfahren angemessene Startpartitionen vorgegeben werden. (Der Autor schlägt die Verwendung des Ergebnisses einer hierarchischen Group-Average-Clusteranalyse vor.) Als Konsequenz der Untersuchung stellen Milligan/Sokol 1980 ein Programm zur Durchführung des Jancey-K-Means-Algorithmus zur Verfügung.

**Programmiersprachliche Formulierungen** (Fortran) von Zerlegungsalgorithmen werden von Anderberg (1973, S. 307ff.), Hartigan (1975, S. 82ff.), Späth (1975, S. 36ff.) und Steinhausen/Langer (1977, S. 109ff.) mitgeteilt. Auch das in manchen Rechenzentren implementierte Programmpaket Clustan von Wishart gestattet die Durchführung von Partitionierungsmethoden (1978, S. 43 ff.). Blashfield/Aldenderfer (1978) diskutieren und evaluieren verschiedene Computerprogramme, die Mengenzersetzungen realisieren.

Für die **Anwendung** sei folgendes Vorgehen empfohlen: Für eine festgelegte Anzahl von Gruppierungen verwende man verschiedene Zerlegungsalgorithmen, soweit vom Verfahren verlangt auch jeweils mit verschiedenen Startpartitionen, und wähle die Aufteilung, welche ein spezifiziertes Kriterium optimiert. (Eine Auswahl nach inhaltlich-theoretischen Gesichtspunkten, z.B. dem sog. ‚Kriterium der Interpretierbarkeit‘ ist nicht angezeigt, da diese zu meist nicht nachvollziehbar ist - für Untersuchungen zu kognitiven Struk-

turierungen besteht zudem die Gefahr der Subjekt-Objekt-Konfundierung; zumindest ist die Mitteilung der hinsichtlich einer explizierten Zielfunktion besten Lösung erforderlich.) Für diese Optimallösung inspiziere man möglichst die Residuen (z.B. bei Klassifikation von Beobachtungswiederholungen die Abweichungen von gruppenspezifischen Mittelwerten für die einzelnen Variablen). Falls gewünscht, wiederhole man dieses Vorgehen für verschiedene Anzahlen von Gruppierungen. In diesem Fall sollte der Verlauf der Werte des Optimalitätskriteriums mitgeteilt werden. Die Auswahl der geeigneten Anzahl von Elementen der Zerlegung erfolgt danach möglichst durch probabilistische oder statistische Evaluation.

*Zur probabilistischen Bewertung der Zerlegbarkeit* von Objektmengen entwickelte parallel zu den Arbeiten von Lingoes/Cooper (1971, siehe auch Lingoes, 1973, S. 176ff., mit Fortran-Programm) und Schultz/Hubert (1973) insbesondere Ling (1971, 1972, 1973) einen exakten graphentheoretischen Ansatz (Kriterium: Verbundenheit ungerichteter Zufallsgraphen.) (Ergänzend sei dazu auf Frank, 1978, und Tinhofer, 1980, verwiesen.) Später wurden Tafeln publiziert (Ling, 1975, Ling/Killough, 1976), die eine partielle Evaluation disjunktiver Gruppierung auf der Basis dieser Theorie gestatten. Nachteil des graphentheoretischen Ansatzes ist die Notwendigkeit der impliziten Verwendung von Schwellenwerten (Cuts) zur Dichotomisierung von Proximitymatrizen (Informationsreduktion). Eine gravierendere Einschränkung der Anwendbarkeit des graphentheoretischen Konnektivitätskriteriums ergibt sich aus der im allgemeinen nicht zutreffenden Annahme, die Ausprägungen der Proximities könnten unabhängig voneinander frei variieren. Dies ist in der Regel nicht der Fall, so sind z.B. Distanzmaße tripelweise wegen der Geltung der Dreiecksungleichung voneinander abhängig. (So argumentieren auch Hubert/Baker, 1977a.)

Deshalb verdient ein Vorschlag von Lee/MacQueen (1980) besondere Beachtung. Zur Prüfung der statistischen Nullhypothese verknüpfen diese Autoren einen K-Means-Algorithmus mit einem Randomization-Test (Edgington, 1969, 1980). Dabei wird der Wert des Optimalitätskriteriums für die empirische Datenmatrix errechnet und mit optimalen Zielfunktionswerten verglichen, die man durch Verfahrensanwendung auf ‚zufällige Datenmatrizen‘ erhält. (Diese werden z.B. durch zufällige Permutation der empirischen Variablenwerte über die Beobachtungswiederholungen generiert.) Obwohl es sehr rechenzeitintensiv ist, sollte dieses Vorgehen als das Verfahren der Wahl für die statistisch konfirmatorische Beurteilung der Zerlegbarkeit von Objektmengen angesehen werden, zumal es ohne weiteres auch auf Datenmatrizen mit nicht numerisch interpretierbaren, z.B. nominalen Merkmalen anwendbar ist.

(Auch Hartigan diskutiert 1975 und 1977 Signifikanztests für Resultate von Zerlegungen (siehe 1975, S. 97ff., S. 135ff.). Seine Überlegungen basieren z.T. auf der statistischen Theorie der (multivariaten) Varianzanalyse. - Lei-

der legen solche Darstellungen dem Anwender völlig inkorrekte Vorgehensweisen nahe, z.B. die Durchführung und ‚statistische Interpretation‘ einer Varianzanalyse nach dem Einsatz einer Partitionierungsprozedur. Davon ist abzusehen, da durch die Verwendung kriteriumoptimierender Zerlegungsverfahren **der** substantielle Bestandteil der statistischen Nullhypothese zerstört wird, nämlich die zufällige Zuordnung zu den Gruppen. Spielarten dieses gravierenden Fehlers finden sich in der inhaltlichen, anwendungsorientierten Literatur sehr häufig, z.B. auch bei der ‚statistischen Beurteilung‘ der Resultate von Techniken der schrittweisen Regression. - Die oben angesprochene Prüfstrategie unter Verwendung von Randomization-Tests unterliegt dieser Kritik nicht, weil sie die Optimierungsverfahren in die Prozedur zur Generierung der Verteilung der statistischen Prüfgröße einbaut.)

Soll nach der empirischen Untersuchung die Übereinstimmung des Resultats von Zerlegungsverfahren mit gegebenen, z.B. theoretisch, a priori konstruierten Partitionen statistisch geprüft werden, so stehen die üblichen nicht-parametrischen Tests zu Kontingenztafeln und die zugehörigen Maße praktischer Signifikanz bzw. Bedeutsamkeit zur Verfügung. Dabei ist wegen der Gewichtungsmöglichkeit in diesem Zusammenhang insbesondere auf Cohens ‚Kappa‘ zu verweisen. (Siehe dazu auch Hubert, 1979.) - Verschiedene (Distanz-)Maße für den Vergleich von Partitionen diskutieren Boorman/Arabie, 1972 und Arabie/Boorman, 1973.

Abschließend bleibt festzustellen, daß die Ausführungen dieses Abschnitts sich bisher auf Vorgehensweisen bezogen, die die Konstruktion lagehomogener Gruppierungen von Objekten zum Ziel haben. Die Suche nach optimalen Mengenerlegungen kann durch geeignete mathematische Formulierung der Bewertungsfunktion aber auch zur Realisierung anderer, komplexerer Zielvorstellungen eingesetzt werden. Nach dem Verweis auf die Analyse von Mischverteilungen (siehe Abschnitt 3.1, Punkt(e)) seien folgende **Ansätze** benannt:

(a) Woodward/Bentler (1979) zerlegen die Menge der Items eines Tests zur Optimierung des Reliabilitätskoeffizienten  $c_r$  in zwei elementenfremde Gruppen.

Ihren einfachen Austauschalgorithmus zur (binären) Zerlegung (Darstellung durch einen Vorzeichenvektor) bezeichnen sie als hoch effektiv. (Deshalb schlagen sie seine Verwendung auch für die Konstruktion lagehomogener Gruppierungen vor.) Von besonderem allgemeinen Interesse ist dazu ihre - eigentlich schon aus anderen Zusammenhängen bekannte - Argumentation: Wenn der Austauschalgorithmus (lediglich) mit einer Wahrscheinlichkeit von  $p$  (z.B.  $p = .2$ ) bei einer zufälligen Ausgangspartition zum globalen Optimum führt (Die Erreichung einer lokalen Optimalkonfiguration ist ohnedies garantiert.), so besitzt er bei  $N$ -maligem Aufruf (z.B.  $N=20$ ) mit verschiedenen zufälligen Startpartitionen eine Wahrscheinlichkeit von  $P = 1-(1-p)^N$  (z.B.  $P = 1-(1-.2)^{20} = .988$ ), daß sich

unter den lokal optimalen Endpartitionen die global beste Zerlegung befindet. (Diese interessante Argumentationsfigur gibt zu der Frage Anlaß, inwieweit es ökonomisch ist, Entwicklungs- und Rechenzeit in kriteriums-spezifische Optimierungsverfahren zu investieren, wenn mit einem Austauschalgorithmus eine generell anwendbare Prozedur vorliegt; allerdings ist vor deren Beantwortung eine jeweilige Abschätzung der Wahrscheinlichkeit  $p$  vonnöten, für bestimmte Problemstellungen zum Optimum zu gelangen.)

(b) Bock (1974, S. 195ff.) diskutiert die Charakterisierung von disjunkten Gruppen durch Hyperebenen. Dabei wird die Zuordnung von Objekten zu Klassen z.B. nicht nach ihrer Entfernung vom Gruppencentroid (bzgl. der numerischen Merkmale) vorgenommen bzw. beurteilt, sondern nach ihrer (senkrechten) Entfernung von einer gruppenspezifischen, im mehrdimensionalen Raum aufgespannten Fläche. Ein Beispiel für diese Problemstellung ist etwa die optimale Aufteilung von Personen in Gruppen, so daß der Variablenraum gruppenspezifisch möglichst einfach (d.h. mit niedriger Dimensionalität) durch Hauptkomponenten repräsentiert werden kann.

(c) Im Anschluß an die unter (b) aufscheinende Grundidee wird hier eine Klasse von Verfahren vorgeschlagen, die mit der Bezeichnung ‚Klassifikationsansatz‘ überschrieben sein soll. Ihre gemeinsame Problemstellung sei wie folgt charakterisiert: Hinsichtlich einer Modalität (der Datenmatrix: Merkmale bzw. Merkmalsträger) sei eine Aufteilung so konstruiert, daß hinsichtlich der (jeweils) anderen Modalität - innerhalb der Gruppen - eine möglichst gute Repräsentation durch ein multivariates Analyseverfahren bzw. -modell erzielt wird.

Ein spezielles Beispiel wäre die unter (b) beschriebene Aufgabenstellung der ‚Klassifikations-Hauptkomponentenanalyse‘.

Ein weiteres Beispiel ist die ‚(multiple) Klassifikationsregression‘, die eine optimale Aufteilung der Beobachtungswiederholungen so sucht, daß innerhalb der Gruppierungen eine möglichst gute Varianzaufklärung einer Kriteriumsvariable durch einen Satz von Prädiktoren erreicht wird. (Dieses Verfahren ‚hinterfragt‘ den allgemeinen Anspruch der ‚Multiple Regression as a General Data Analytic System‘ (Cohen, 1968). - Den Spezialfall nur eines Prädiktors mit homogenen (identischen) Steigungskoeffizienten für die (nicht essentielle) Restriktion auf zwei Gruppen untersucht bereits Mustonen (1978).)

Weiterhin generiert dieser allgemeine, als Heuristik dienende Ansatz (aus-schöpfend) z.B. die Verfahren der ‚Klassifikationsanalyse kanonischer Korrelation‘, der ‚Klassifikations-Mengenzerlegungsanalyse‘ und der ‚Klassifikationsanalyse hierarchischer Clusterung‘. Schließlich können sogar Verfahren der ‚Klassifikations-Varianzanalyse‘ und der ‚Klassifikations-Diskriminanzanalyse‘ einbezogen werden.

Offensichtlich steht der hier entwickelte allgemeine Klassifikationsansatz in engster Beziehung zur klassischen Analyse latenter Klassen (z.B. Lazarsfeld/Henry, 1968); er könnte in der Anwendung vornehmlich der Homogenitätsprüfung von Modellapproximationen zu Samples dienen; der methodologischen Literatur bleibt die Bereitstellung von Überprüfungsverfahren vorbehalten.

Für die praktische Umsetzung der hochkomplexen Aufgabenstellungen der Analyse von Mengenerlegungen, z.B. des Klassifikationsansatzes, wird man zur Optimierung der (jeweils zu spezifizierenden) Zielfunktion vorläufig auf den Einsatz der relativ generellen Austauschverfahren zurückgreifen (siehe z.B. Banfield/Basill, 1977, mit Fortran-Programm) und die beste Lösung der Anwendung eines Transferalgorithmus auf verschiedene Startpartitionen auswählen. - Im Anwendungsfall sind exploratorische Bemühungen (sensu Tukey, 1977), z.B. in Form einer genauen Residueninspektion, und die Entwicklung und Durchführung statistischer Prüfungen, z.B. durch angemessen konstruierte Randomization-Tests, unverzichtbar, da sonst die Gefahr der Interpretation von Artefakten zu groß würde. Dabei ist der rechenzeitintensive Einsatz von Computern zur genauen Prüfung von Verfahrensergebnissen bzw. Modellen nicht nur notwendig, sondern auch sinnvoll und wünschenswert.

### 3.3 Hierarchische Clusteranalysen

#### **3.3.1 Einordnung und Charakteristik**

Hierarchische Verfahren sind in der Psychologie die am weitesten verbreiteten Methoden aus dem Bereich der Clusteranalyse (Blashfield/Aldenderfer, 1978). Der Boom in der Methodendiskussion und der inhaltlich orientierten Anwendung geht auf den schon klassischen Artikel ‚Hierarchical Clustering Schemes‘ von Stephen C. Johnson (1967) zurück (Blashfield, 1980). Dabei hatte der Autor lediglich Verfahren der ‚numerischen Taxonomie‘ aus der Biologie übernommen (Sokal/Sneath, 1963) (und parallel zu Hartigan (1967) die Charakterisierbarkeit der Verfahrensergebnisse durch die ultrametrische Ungleichung (siehe Abschnitt 2) aufgezeigt). Die breite Aufnahme der für die Psychologie neuen Methodenklasse ist wohl auch weniger auf die allgemeine Angemessenheit der Verfahren zur strukturellen Repräsentation von Proximitymatrizen zurückzuführen, vielmehr traf Johnsons Brückenschlag auf einen Boden, der durch das stark aufkeimende ‚Unbehagen in der Faktorenanalyse‘ (Kallina, 1967) aufbereitet war und damit die Suche nach alternativen Analysemethoden förderte.

Ausgangssituation für den Einsatz von Verfahren der hierarchischen Clusteranalyse (HCA) ist das Vorliegen einer Proximitymatrix (von Ähnlichkeitsma-

Ben  $s_{kl}$  oder Unähnlichkeitsmaßen bzw. Distanzen  $d_{kl}$ ) zu einer bestimmten Objektmenge  $O$  mit  $n$  Elementen. Aufgabe von Verfahren der HCA ist die Repräsentation der Proximities durch die Angabe einer geordneten Folge von Zerlegungen  $Z(O)_0, Z(O)_1, \dots, Z(O)_h, \dots, Z(O)_{n-1}$ , wobei  $Z(O)_0$  die feinste und  $Z(O)_n$ , die größte Partition zu  $O$  ist. (Zugang sensu Ward, 1963; zur Definition der hier verwendeten Begriffe siehe Abschnitt 2 und 3.2.)

Ferner soll für alle  $h$  mit  $0 \leq h \leq n-2$  gelten, daß  $Z(O)_{h+1}$  eine strikte Vergrößerung von  $Z(O)_h$  ist, d.h. (genau) eine Gruppierung aus  $Z(O)_{h+1}$  ist durch Vereinigung von (mindestens) zwei Gruppierungen aus  $Z(O)_h$  entstanden. Umgekehrt heißt  $Z(O)_h$  auch Verfeinerung von  $Z(O)_{h+1}$ . Entnimmt man also aus beliebigen zwei Zerlegungen der geordneten Folge jeweils eine Gruppierung, so sind diese entweder elementenfremd oder eine Gruppierung ist Untermenge der anderen. Dieser Sachverhalt charakterisiert ein hierarchisches (Mengen)System, siehe z.B. (Bock, 1974, S. 360.)

Geordnete Folgen von Zerlegungen können explizit oder grafisch durch sog. Dendrogramme angegeben werden; Beispiele finden sich in der unten aufgeführten Literatur. Diese veranschaulichen die sukzessive Vergrößerung der Partitionen durch Fusion von Gruppierungen bzw. umgekehrt die sukzessive Verfeinerung durch Separation von Gruppierungen. Außerdem eröffnen sie die Möglichkeit, den Zusammenlegungen bzw. Trennungen bestimmte (numerische) Niveaus zuzuordnen; aus ihnen können für alle Paare von Objekten aus  $O$ , über die Höhe der Knoten, bei der die Objekte durch Kantenfolgen zusammengeführt werden, Proximities  ${}_u s_{kl}$  bzw.  ${}_u d_{kl}$  bestimmt werden. Ist man bei der Analyse von Unähnlichkeitsmaßen oder Distanzen ausgegangen, so erfüllen die  ${}_u d_{kl}$  die ultrametrische Ungleichung. (Zur Definition siehe Abschnitt 2; im Fall der Analyse von Ähnlichkeitsmaßen erfüllen die  ${}_u s_{kl}$  die der Ultrametrik äquivalente Eigenschaft: für alle  $o_k, o_l, o, EO$  gilt  ${}_u s_{ks} \geq 2 \min \{ {}_u s_{kl}, {}_u s_{ls} \}$ ); hier wird von einzelnen Verfahren abgesehen, die sog. Inversionen zulassen (siehe Eckes/Roßbach, 1980, S. 73), und deshalb verschiedentlich zu nicht interpretierbaren Resultaten führen (ein empirisches Beispiel findet man bei Steinhausen/Langer, 1977, S. 92f.).)

Da Dendrogramme und Matrizen ultrametrischer Proximities einander eindeutig zugeordnet sind, können hierarchische Clusteranalysen als Transformationsverfahren von empirischen Proximitymatrizen in Ultrametrien charakterisiert werden. Dies gibt Anlaß zu einer wichtigen Reformulierung der Aufgabenstellung von HCA-Methoden: zu einer Matrix, z.B. von Unähnlichkeitsmaßen  $[d_{kl}]$ , sollen sie eine ultrametrische Matrix, z.B.  $[{}_u d_{kl}]$ , so konstruieren, daß hinsichtlich einer spezifizierten Zielfunktion  $Z = f([d_{kl}], [{}_u d_{kl}])$  die Ultrametrik eine optimale Repräsentation der (empirisch) gegebenen Proximitydaten darstellt. - Eine Liste solcher Optimalitätsmaße gibt Cormack (1973, S. 338); diese wurden zwar verschiedentlich zur nachträglichen Evaluation der Resultate von HCA-Verfahren vorgeschlagen, (so nennt Johnson (1967) z.B.

die Rangkorrelation), ihr konstruktiver Einsatz zur Generierung von Lösungen eines (diskreten) Optimierungsproblems wurde bisher jedoch allzu stark vernachlässigt. Die Methoden der HCA erhalten dann einen lediglich heuristischen Charakter, d.h. die Resultate der Verfahrensanwendungen führen zu verschiedenen Vorschlägen zur Lösung des Repräsentationsproblems. In der Forschungspraxis äußert sich dies manchmal in einer Verfahrensselektion nach persönlicher Vorliebe und in einer schwer nachvollziehbaren Auswahl ihrer Resultate nach einem Kriterium der ‚theoretischen Angemessenheit‘ bzw. ‚Interpretierbarkeit‘. (Ein Vorschlag für ein begründeteres Vorgehen wird in Abschnitt 3.3.4 dargestellt. - Es gibt aber auch Autoren, die den konstruktiven Optimierungsansatz verfolgen: Hartigan (1967), Kruskal/Carroll (1969), Carroll/Pruzansky (1980).)

Systematische übersichten und zahlreiche Beispiele zu den (heuristischen) Methoden der HCA geben: Anderberg (1973, S. 131ff.), Bock (1974, S. 356ff.), Eckes/Roßbach (1980, S. 63ff.), Hartigan (1975, S. 191ff.), Jardine/Sibson (1971, S. 45ff.), Opitz (1980, S. 96ff.), Späth (1975, S. 147ff.), Steinhäusen/Langer (1977, S. 73 ff.) und Vogel (1975, S. 249ff.).

Zur Konstruktion von geordneten Folgen von Partitionen der Objektmenge gibt es zwei übliche Gruppen von Vorgehensweisen: Die agglomerativen Verfahren beginnen mit der feinsten Partition und legen sukzessive Gruppierungen nach einem Kriterium der maximalen Ähnlichkeit bzw. minimalen Distanz zusammen. Die subdivisiven Verfahren beginnen mit der größten Partition und zerlegen sukzessive die Gruppierungen nach einem Kriterium der minimalen Ähnlichkeit bzw. maximalen Distanz. Es ist aber durchaus nicht zwingend, den Analyseprozeß bei den trivialen Partitionen zu beginnen, vielmehr ist es mindestens ebenso sinnvoll, zunächst die Konstruktion einer optimalen Mengenerlegung (siehe Abschnitt 3.2) mit einer mittleren Gruppenzahl, z.B. von ca.  $n/2$ , vorzunehmen, um dann, von dieser Partition ausgehend, die Folge der Zerlegungen agglomerativ ‚nach oben‘ und subdivisiv ‚nach unten‘ zu vervollständigen. Für die Praxis böte sich gerade ein solches Vorgehen an. Da sehr häufig nicht das Gesamtdendrogramm, sondern ohnedies eine Partition ‚mittleren Niveaus‘ zur Interpretation herangezogen wird, liegt es zumindest für diese Fälle nahe, die Bemühung um Optimierung (der Abbildungsstreuung) auch in diesem Bereich anzusetzen.

Da für alle gängigen HCA-Verfahren die Ausgangsdaten (Proximities) Ähnlichkeiten bzw. Distanzen zwischen einzelnen Objekten angeben, ist die definitorische Festlegung dessen, was unter Ähnlichkeit bzw. Distanz von Gruppierungen anzusehen ist, zentrales methodengenerierendes Prinzip. Betrachtet man z.B. Distanzen, so gilt es folgende Kriteriumsfunktion zu spezifizieren:

$$(F.1) \quad d(G_p, G_q) = \underset{o_k \in G_p, o_l \in G_q}{\text{Funktion}} \{d(o_k, o_l)\}$$

Setzt man z.B. fest, daß man agglomerativ vorgehen will, und daß für Methode  $M_1$ : ‚Funktion‘  $\equiv$ : ‚Min‘, für  $M_2$ : ‚Funktion‘  $\equiv$ : ‚Max‘ und für  $M_3$ : ‚Funktion‘  $\equiv$ : ‚arithmetisches Mittel‘, so hat man drei prominente HCA-Verfahren konstruiert:

- $M_1$  heißt ‚Single Linkage‘ oder ‚Minimum Methode‘; für den Fall der Fusion von  $G_p$  und  $G_q$  garantiert sie, daß zwei Objekte (je eines aus  $G_p$  und  $G_q$ ) mindestens eine Distanz von  $d(G_p, G_q)$  aufweisen.

- $M_2$  heißt ‚Complete Linkage‘ oder ‚Maximum Methode‘; sie garantiert, daß in der neuen Gruppierung  $G_p \cup G_q$  keine paarweise Distanz zwischen Objekten größer als  $d(G_p, G_q)$  ist, weshalb sie auch manchmal ‚Diameter Methode‘ genannt wird.

- $M_3$  heißt ‚Group-Average Methode‘.

Dieser Darstellungsmodus verdeutlicht, welche Vielfalt von heuristischen Verfahren mit einfachen Mitteln konstruierbar ist: es sind prinzipiell unendlich viele. Damit wird wiederum die Notwendigkeit des konsequenten konstruktiven Einsatzes von Optimierungsverfahren unterstrichen.

Andererseits kann im Zusammenhang mit (F.1) auch leicht erklärt werden, warum agglomerative Verfahren sich gegenüber subdivisiven Verfahren so großer Beliebtheit erfreuen: Sie machen nicht so viel Arbeit bzw. sind leicht von Hand durchführbar ( $n < 20$ ). Während bei fusionierendem Vorgehen z.B. auf unterster Stufe (feinste Partition) lediglich  $n(n-1)/2$  Werte  $d(G_p, G_q)$  (die alle noch gleich  $d(o_k, o_l)$  sind) untersucht werden müssen, sind es beim separierenden Vorgehen auf oberster Stufe (größte Partition)  $2^{n-1} - 1$  mögliche (binäre) Zerlegungen mit zugehörigen Werten  $d(G_p, G_q)$ . Diese ‚Ökonomieüberlegung‘ spricht aber keineswegs von vornherein für die agglomerativen Verfahren, da diese auf niedriger Stufe aufgrund relativ weniger Werte durch die Zusammenlegungen Vorentscheidungen treffen, die in späteren Stadien des Analyseprozesses u.U. ‚ungünstige‘ Fusionen erzwingen. Man kann aber (wiederum, s. o.) davon ausgehen, daß Anwender häufig an der Interpretation relativ weniger großer Gruppen interessiert sind. Deshalb sollte deren Zusammenstellung auch möglichst sorgfältig erfolgen.

### **3.3.2 Agglomerative Verfahren**

Speziell für die kompakte tabellarische Darstellung der verbreiteten agglomerativen Prozeduren wird in der Literatur eine Schreibweise tradiert, die auf Lance/Williams (z.B. 1967) zurückgeht. Wurden auf einer bestimmten Stufe der Analyse zwei Gruppierungen  $G_p$  und  $G_q$  aufgrund z.B. des Kriteriums der minimalen Distanz bzw. der maximalen Ähnlichkeit für die Aggregation aus-

Tabelle: Agglomerative hierarchische Clusteranalysen.

Name	Parameter		Eigenschaften	
	$\alpha_p$	$\alpha_q$	$\beta$	$\gamma$
1 Single Linkage	1/2	1/2	0	-1/2
2 Complete Linkage	1/2	1/2	0	1/2
3 Group Average	$n_p/(n_p+n_q)$	$n_q/(n_p+n_q)$	0	0
4 Weighted Average	1/2	1/2	0	0
5 Centroid	$n_q/(n_p+n_q)$	$n_p/(n_p+n_q)$	$-n_p n_q / (n_p + n_q)^2$	0
6 Median	1/2	1/2	-1/4	0
7 Minimum Variance	$(n_p + n_r) / (n_p + n_q + n_r)$	$(n_q + n_r) / (n_p + n_q + n_r)$	$-n_r / (n_p + n_q + n_r)$	0
8 Flexible	$(1 - \beta) / 2$	$(1 + \beta) / 2$	$\beta (< 1)$	0

Gebräuchliche alternative Bezeichnungen:

- Zu 1: Nearest Neighbor Method, Connectedness Method, Minimum Method
- Zu 2: Furthest Neighbor Method, Diameter Method, Maximum Method
- Zu 3: Unweighted Pair-Group Method Using Arithmetic Averages (UPGMA)
- Zu 4: Weighted Pair-Group Method Using Arithmetic Averages (WPGMA)
- Zu 5: Unweighted Pair-Group Centroid Method (UPGMC)
- Zu 6: Weighted Pair-Group Centroid Method (WPGMC)
- Zu 7: Error Sum of Squares Method (HGROUP)

gewählt, so ist zunächst zu bestimmen, welche Distanz die neue Gruppierung  $G_p \cup G_q$  zu den anderen Gruppen  $G_r$  haben soll. Durch Spezifikation der Parameter  $\alpha_p$ ,  $\alpha_q$ ,  $\beta$  und  $\gamma$  der folgenden Rekursionsformel ergeben sich dann verschiedene agglomerative HCA-Verfahren:

$$(F.2) \quad d(G_r, G_p \cup G_q) = \alpha_p d_{rp} + \alpha_q d_{rq} + \beta d_{pq} + \gamma |d_{rp} - d_{rq}|$$

Dabei wurde z.B.  $d(G_r, G_p) \equiv: d_{rp}$  verwendet.

In der Tabelle stehen  $n_p$ ,  $n_q$ , und  $n_r$  für die Anzahl der Elemente in den Gruppen  $G_p$ ,  $G_q$  und  $G_r$ . (Ähnliche Übersichten geben Milligan, 1979 (S. 344) und Steinhausen/Langer, 1977 (S. 77). Referenzen zu den einzelnen Verfahren gibt Cormack, 1971 (S. 331). - Für das ‚Single Linkage‘ ergibt sich tatsächlich  $d(G_r, G, UG_r) = \min(d_{rp}, d_{rq})$ , da man für beliebige reelle Zahlen  $x, y$  schreiben kann:  $\min(x, y) = (x + y - |x - y|) / 2$ .)

Eine an der Rekursionsformel orientierte Übersicht der agglomerativen Verfahren geben Eckes/Roßbach (1980, S. 67f.), sowie Steinhausen/Langer (1977, S. 76ff.). (Für die von Lance/Williams 1967 aufgrund der Rekursionsformel entwickelte ‚Flexible Strategie‘ schlagen die Autoren selbst einen Wert von  $\beta = -.25$  vor. Steinhausen/Langer (1977, S. 77) empfehlen das Intervall  $-.2 \leq \beta \leq -.4$ .)

Milligan (1979) verwendet die Rekursionsformel, um für bestimmte Parameterwerte  $\alpha_p$ ,  $\alpha_q$ ,  $\beta$  und  $\gamma$  nachzuweisen, daß die zugehörigen Verfahren die Eigenschaft haben, monotone (ultrametrische) Hierarchien (ohne Inversionen) zu liefern. Danach ist diese Eigenschaft gegeben, wenn  $\alpha_p + \alpha_q + \beta \geq 1$  mit  $\alpha_p, \alpha_q, \gamma \geq 0$  oder wenn  $\gamma < 0$  mit  $|\gamma| \leq \alpha_p, \alpha_q$  gilt.

Die Beachtung der Verfahrenseigenschaft, invariante Ergebnisse (bezüglich der Folge der Zerlegungen) bei monotoner Transformation der Proximities zu liefern (Johnson, 1967; Hubert, 1973), hat ihre Wurzeln in der nicht-metrischen mehrdimensionalen Skalierung (siehe z.B. Ahrens, 1974, S. 163ff.). Die Methoden des ‚Single-Linkage‘ und des ‚Complete-Linkage‘ erfüllen dieses Kriterium, da sie bei der Ermittlung der Distanzen zwischen einer ‚neuen‘, durch Fusion entstandenen Gruppierung  $G_p \cup G_q$  und den jeweils ‚anderen‘ Gruppierungen  $G_r$  lediglich ‚Kleiner-‘ bzw. ‚Größer-Abfragen‘ vornehmen, während durch alle anderen Verfahren ‚echte‘ arithmetische Rechenoperationen durchgeführt werden.

Fraglich ist allerdings, ob man dem Kriterium der ‚monotonen Invarianz‘ - wie bisher - eine besondere Dignität zuschreiben sollte. Es besagt zwar, daß die Verfahren zur Analyse nur die in den Daten vorhandene Ranginformation nutzen; dies bedeutet aber nicht, daß die Ergebnisse hinsichtlich der Repräsentation der (empirischen) Ranginformation auch optimal sind. Im Gegenteil, da ‚Single-Linkage‘ und ‚Complete-Linkage‘ nur die Information der jeweiligen

Extremwerte in den nächsten Analyseschritt transferieren, dürften die metrisch arbeitenden Verfahren in der Regel, auch bezogen auf nicht-metrische Kriteriumsmaße, wie z.B. Kruskals ‚Stress‘ oder Guttman's Maß(e) der schwachen Monotonie, Ultrametrien generieren, die die Ranginformation in den Proximitydaten besser abbilden. (Nebenbei sei auf die einfache Möglichkeit hingewiesen, allen HCA-Verfahren die Eigenschaft der Invarianz bei monotoner Transformation dadurch zuzueignen, daß man vor Verfahrensanwendung die Proximities rangiert, um dann die Ränge metrisch zu behandeln. Ein solches Vorgehen dürfte aber wohl als krude zu bezeichnen sein.)

Charakteristisch für die Methoden des ‚Single-Linkage‘ und des ‚Complete-Linkage‘ sind ihre **Verzerrungseigenschaften** („Spate-Distortion“, Lance/Williams, 1967; Hubert/Schultz, 1975). Das ‚Single-Linkage‘ neigt dazu, Objekte bzw. Gruppierungen sukzessive aneinanderzureihen (Kontraktion, ‚chaining-effect‘). Es tendiert zur Ausbildung einzelner großer Gruppen mit Durchmessern, die größer sein können als die Distanzen zwischen Objekten aus verschiedenen Gruppierungen. Das ‚Complete-Linkage‘ neigt zur Ausbildung gleichgroßer Gruppen; das zugehörige Dendrogramm macht zumeist einen ‚stark strukturierten Eindruck‘; damit gibt es Separationen von Objekten bzw. Gruppierungen vor, die relativ eng zusammenliegen können (Dilatation). - Diese Hinweise auf mögliche Artefaktquellen in den Verfahren gehen auf Lance/Williams (1967) zurück; sie wurden z.B. von Cormack (1971, S. 331) als zuwenig formalisiert kritisiert, Baker/Hubert (1975) aber konnten die Angemessenheit der Verfahrenscharakterisierung stützen, indem sie in einer Simulationsstudie die durch die Methoden produzierten Typen von Partitionen (Verteilungen der Gruppengrößen) untersuchten (siehe auch Hubert/Schultz, 1975).

Zur Beurteilung der **Leistung** der **Rekursionsformel** können die folgenden Gesichtspunkte herangezogen werden:

- Sie leistet eine exakte und kompakte Systematisierung der (bis 1967) bekanntesten agglomerativen HCA-Verfahren und entwickelt daraus eine ganze Klasse neuer heuristischer Methoden („Flexible Strategie“). Dieses beinhaltet allerdings auch die Gefahr der Einschränkung des Blickwinkels aufgrund des Eindrucks, alle denkbaren Methoden müßten durch die Parameter der Rekursionsformel beschreibbar sein.

- Sie leistet eine ökonomische prozeßbezogene Beschreibung der Analysen, indem sie verdeutlicht, daß während der Durchführung der Verfahren nicht ständig auf die ursprüngliche Matrix der Proximities  $d(o_k, o_l)$  zurückgegriffen werden muß; ausreichend ist die Untersuchung der Proximities zwischen den Gruppen  $d(G_p, G_q)$ , deren Anzahl geringer ist (Informationsreduktion). Dieser Gesichtspunkt war früher bei der Realisierung der Algorithmen auf elektron-

schen Rechenanlagen von Belang (Speicherplatzbedarf), seine Beachtung ist aber jetzt, außer bei Problemen mit sehr großen Anzahlen von Objekten (z.B.  $n > 500$ ), nicht mehr zeitgemäß.

Es sollen nun **weitere agglomerative HCA-Verfahren** angesprochen werden, die nicht mehr mit Hilfe der Rekursionsformel charakterisierbar sind; sie verlangen während des gesamten Analyseprozesses die Untersuchung der ursprünglichen empirischen Proximities  $d(o_k, o_l)$ . Die hier benannten Methoden haben die Eigenschaft, daß ihre Ergebnisse bei monotoner Transformation der Proximities invariant bleiben; dabei rechnen sie nicht nur in den (jeweiligen) Extremwerten, wodurch die unangenehmen Verzerrungseigenschaften des ‚Single-‘ und ‚Complete-Linkage‘ teilweise vermieden werden können:

(a) D'Andrade (1978) schlägt ein Verfahren vor, das auf den jeweiligen Stufen die Gruppierungen zusammenlegt, die nach dem, aus der nicht-parametrischen Statistik entlehnten, Kriterium der Mann-Whitney-U-Statistik maximal ähnlich bzw. minimal distant sind. Die Methode verwendet zur Fusionsentscheidung lediglich Auszählungen der Ergebnisse von Größenvergleichen für Proximities von Objektpaaren aus verschiedenen Gruppen (sie führt deshalb zu Dendrogrammen, die nicht numerisch evaluiert sind, sog. ‚bare trees‘ (Boorman/Olivier, 1973)).

(b) Durch Setzung von ‚Funktion‘  $\equiv$  : ‚Median‘ in (F.1) entsteht ein Verfahren, das die Gruppen nach dem Kriterium des minimalen Medians der Distanzen von Objekten aus zwei verschiedenen Gruppen zusammenfaßt. (Diese Methode darf nicht mit der geometrisch orientierten ‚Median Method‘ (Nr. 6 der Übersicht) verwechselt werden.) (Weitere Verfahren ergeben sich durch Einsetzen anderer, insbesondere robuster Maße der zentralen Tendenz der Proximities von Objekten aus (zwei) verschiedenen Gruppen.) Speziell die Verwendung des Medians wurde, wie Johnson (1967) mitteilt, von J. D. Carroll (mündlich) vorgeschlagen.

(c) Ergänzend ist an dieser Stelle auf ein Verfahren von Hubert (1973, S. 55), die ‚Objective Function Method‘ und auf die Klasse der ‚r-Diameter-Hierarchical-Clusterings‘ hinzuweisen (Hubert/Schultz, 1975; Hubert/Baker, 1977). Diese Methoden sind zwar ‚monoton invariant‘, aber nicht ‚konservativ‘, d.h. sie zeigen in unterschiedlichem Maße Verzerrungen.

Die unter (a) bis (c) angesprochenen heuristischen agglomerativen HCA-Verfahren zeigen zwar gegenüber ‚Single-‘ und ‚Complete-Linkage‘, wie von D'Andrade (1978) und Hubert/Baker (1977) für ihre Methoden ausgewiesen wird, die besseren Goodness-of-Fit Leistungen hinsichtlich des nicht-metrischen Goodman-Kruskal  $\gamma$ , es kann aber (noch) nicht allgemein garantiert werden, daß die von den Verfahren generierten Resultate monoton hierarchisch, d.h. ultrametrisch sind.

### 3.3.3 Subdivisive Verfahren

Für die Klasse der ‚absteigenden‘ HCA-Methoden, die von der größten Partition  $Z(0) = 0$  ausgehen und die Zerlegungen sukzessive verfeinern, kann generell ausgesagt werden, daß konzeptuell alle Ansätze der Mengenerlegung (Abschnitt 3.2) und der agglomerativen Verfahren (s.o.) als Heuristiken zur Konstruktion divisiver Methoden herangezogen werden können. Für zusammenfassende Darstellungen, wie der hier vorliegenden ist es daher verständlich, wenn der Diskussion zerlegender hierarchischer Methoden weniger Raum gegeben wird. Die geringe Verbreitung in der Anwendung ist darüber hinaus leicht aus den gewöhnlich immensen Anzahlproblemen der (binären) Partitionierung erklärt. Solange aber kriteriumsgeleitete Optimierungsverfahren zur Approximation von Ultrametrikern an empirische Proximitymatrizen nicht allgemein verfügbar sind, sollte insbesondere den heuristischen Verfahren der sukzessiven Zerlegung erhöhte Beachtung zukommen. Dieser Empfehlung soll mit der folgenden Argumentation (nochmals) Nachdruck verliehen werden: Während agglomerative HCA-Verfahren zunächst in den kleinen Distanzen arbeiten und (Vor-)Entscheidungen treffen, gehen divisive Methoden anfangs von der Untersuchung der großen Distanzen bzw. geringen Ähnlichkeiten aus; es sind aber gerade die großen Entfernungen, welche für die Global- oder Gesamtstruktur einer Objektmenge besonders prägend sind. Die Geltung dieser Aussage wird durch eine Untersuchung von Graef/Spence (1976, zitiert bei Kruskal, 1977) gestützt. Die Autoren konnten durch die Anwendung von MDS-Algorithmen nachweisen, daß die Streichung kleiner Distanzen aus der Proximitymatrix auf die Struktur von Punkten im Raum erwartungsgemäß weit geringer deformierend wirkt, als die Streichung großer Distanzen. Deshalb sollten Investitionen in die Analyse der größeren Unähnlichkeiten lohnend sein, zumal die zunehmende Verbreitung schneller Computer solchen intensiven Bemühungen entgegenkommt.

Im Einzelnen sei ergänzend auf folgende Ansätze der sukzessiven Zerlegung hingewiesen:

- (a) Hubert (1973) stellt (neben einem weiteren Verfahren) subdivisive Pendants zum agglomerativen ‚Single-‘ und ‚Complete-Linkage‘ zur Verfügung. Diese Methoden liefern ebenfalls invariante Resultate bei monotoner Transformation der Proximitymatrizen und sind zudem ökonomisch durchführbar.
- (b) Ein zum agglomerativen ‚Minimum Variante‘ Verfahren nach Ward (1963) (Methode 7 der Übersicht in 3.3.2) gehörendes subdivisives Pendant haben Edwards/Cavalli-Sforza (1965) vorgeschlagen. Ihm wird wegen der benötigten extremen Rechenzeiten z.B. von Eckes/Roßbach lediglich „illustrativer Wert“ (1980, S. 80) zugeeignet. Diese Einschätzung muß aber wohl revidiert werden, nachdem Herden/Steinhausen (1979 mit Fortran-Pro-

gramm) einen effektiven Branch-and-Bound Algorithmus zur Berechnung des globalen Minimums des Varianzkriteriums angeben.

- (c) Für die (in Abschnitt 3.3.2 angesprochene) Festlegung ‚Funktion‘=: ‚Median‘, bei der im subdivisiven Fall der Median der paarweisen Distanzen von (zwei) Gruppen für die Aufteilung maximiert wird, können - wie auch für andere komplexe Kriteriumsfunktionen - Austauschalgorithmen verwendet werden. (Diesem Vorgehen folgt Späth (1976) bei Vorliegen von Datenmatrizen  $[x_{ij}]$  für die Minimierung des  $L_1$ -Kriteriums: Summe der absoluten Abweichungen vom Median. Siehe auch Späth (1977) und Banfield/Basill (1977), sowie Späth/Müller (1979) für weitere Anwendungen (jeweils mit Fortran-Programmen).)
- (d) Eine interessante Beobachtung führte McQuitty (z.B. 1968) zur Entwicklung seiner ‚Iterative Intercolumnar Correlational Analysis‘: Korreliert man paarweise die Zeilen einer Korrelationsmatrix  $K_1$  zur Herstellung einer Korrelationsmatrix  $K_2$  und iteriert dieses Vorgehen, so tritt gewöhnlich (‚usually‘, S. 466) ein Konvergenzphänomen ein:  $K_N = K_{N+1}$ . Die Matrix  $K_N$  hat dann nur noch Eintragungen von +1 und -1 und ist einer Zerlegung der Menge der Variablen eineindeutig zugeordnet. Wendet man dieses Vorgehen sukzessive auf die entstandenden Untergruppen bzw. deren zugehörige Untermatrizen an, so entsteht im (zumeist) erfolgreichen Fall eine Folge von (binären) Zerlegungen. Das Verfahren kann somit als heuristische subdivisive HCA angesehen werden. Es hat zumindest folgende Schwächen :
- (1) Auf jeder Stufe wird die Zerlegung ohne jedes explizite Optimierungskriterium vorgenommen. Diese Eigenschaft teilt die Methode mit anderen Zerlegungsprozeduren.
  - (2) Es ist fraglich, ob die Produkt-Moment-Korrelation wegen ihrer Invarianz bei linearen Transformationen als Maß für die Übereinstimmung von Zeilen von Korrelationsmatrizen geeignet ist.
  - (3) Die Konvergenzeigenschaft des Vorgehens konnte, trotz der Begründungsbemühung von McQuitty/Clark (1968) bisher nicht bewiesen werden.

Dazu sei hier mitgeteilt, daß (nach eigenen Untersuchungen) auch bei Verwendung verschiedener Distanzmaße zur iterativen Analyse dieser Art, nach einigen Schritten (mit Normierung) Konvergenz in dem Sinn eintritt, daß die resultierende Distanzmatrix eine Verteilung hoher gegenüber sehr niedrigen Eintragungen enthält, die einer Zerlegung zugeordnet ist. (Ergänzend sei auf eine vergleichbare Technik von Bonner (1964) verwiesen, der Ähnlichkeitsmaße verwendete.) - Solche Proximitymaße von Proximities und ihre Leistungen zur Generierung disjunkter Hierarchien sollten verstärkt formal und im Rahmen von (vergleichenden) Evaluationsstudien untersucht werden. Die inneren Zusammenhänge dieser Vorgehensweise und ihre Effizienz für die Anwendung bedürfen der Klärung.

### 3.3.4 Evaluation, Anwendung und Weiterentwicklungen

Evaluationsstudien zu HCA-Verfahren liegen m. W. ausschließlich für deren agglomerative Varianten vor. Zwei Typen von Studien lassen sich grob unterscheiden :

- (a) Tests nach dem **Modell der Mischverteilungen** gehen von einer Gesamtpopulation aus, die sich aus mehreren Subpopulationen mit unterschiedlichen Verteilungstypen, Mittelwert(svektor)en und/oder Kovarianzmatrizen zusammensetzt. Aus der Gesamtpopulation werden Stichproben generiert und den HCA-Verfahren unterworfen, zur Leistungsbewertung wird abschließend die Herkunft der Objekte (Punkte) aus den jeweiligen Subpopulationen mit ihrer Clusterzuordnung durch die Methoden über Kappa- oder Rand-Statistiken verglichen (Blashfield, 1976, Edelbrock, 1979, Edelbrock/McLaughlin, 1980, McIntyre/Blashfield, Milligan, 1980).

Bei insgesamt guten Zuordnungsleistungen sind die Verfahren (im Sinn einer Interaktion) differentiell sensitiv für verschiedene, durch das Populationsmodell spezifizierte Ausgangssituationen. Es werden also keine generellen Empfehlungen ausgesprochen. (Nach meinem Eindruck ergibt eine Übersicht leichte Vorteile für die ‚Average‘-Methoden und die ‚Minimum-Variante‘-Technik.) Relativ einhellig dagegen sind die Warnungen vor der Betrachtung nur einer Lösung, da die Verfahren in der Regel sehr verschiedene Resultate erbringen, die es jeweils genau zu untersuchen gilt.

(Für diesen Typ von Evaluationsstudien lassen sich die Leistungen der HCA-Verfahren mit denen der Mengenerlegungsmethoden vergleichen. Dabei sind die Ergebnisse der für diesen Zweck speziell konstruierten Partitionierungsalgorithmen (z.B. aus der K-Means-Gruppe, Milligan, 1980) erwartungsgemäß besser, wenn ihnen nur plausible Startpartitionen (z.B. HCA-Verfahren als Heuristik) zur Verfügung gestellt werden.)

- (b) Untersuchungen zur Anfälligkeit von HCA-Verfahren bei **Vorliegen von Fehlern in den Daten** gehen von der Konstruktion ‚wahrer‘ metrischer oder ultrametrischer Proximitymatrizen aus. Ultrametrien werden dann mit aus verschiedenen Verteilungen stammenden zufälligen Fehlern überlagert und den HCA-Verfahren unterworfen. Zur Leistungsbewertung wird abschließend die Ranginformation in der ursprünglichen Metrik oder Ultrametrik mit der Ranginformation in den durch die Verfahren generierten Ultrametrien verglichen; dazu verwendet man Kendalls Tau oder Goodman-Kruskals Gamma (Baker, 1974; Cunningham/Ogilvie, 1972; D'Andrade, 1978; Hubert/Baker, 1977). Auch in diesen Untersuchungen ergibt sich eine Interaktion von Ausgangssituation und Methoden: Sind die ‚wahren‘ (Ultra-)Metriken ‚verkettet‘, so wird dies vom ‚Single-Linkage‘ besser reproduziert; ansonsten stimmen die Ergebnisse des ‚Complete-Linkage‘, vor allem aber die ‚mittleren‘ Verfahren, z.B. ‚Group-Average‘

und ‚Minimum Variante‘, deutlich besser mit den ursprünglichen Daten überein.

Da dem Anwender in der Regel nicht bekannt ist, inwieweit ‚Verkettungen‘ in seinen Proximitydaten vorliegen, bleibt ihm (wiederum) nur die Durchführung mehrerer HCA-Verfahren und die genaue Untersuchung ihrer Ergebnisse.

Evaluationsstudien führen also nicht zu generellen Empfehlungen. Vielmehr scheinen die heuristischen HCA-Verfahren für verschiedene Aspekte oder Merkmale der Daten sensibel zu sein, bzw. scheinen durch ihre Darstellungsergebnisse unterschiedliche ‚Aspekte‘ oder Merkmale der Daten hervorzuheben oder ihnen aufzuprägen.

Die für die ökonomische Anwendung von HCA-Verfahren notwendigen *Programme* wurden überwiegend für die agglomerativen Varianten publiziert, z.B. Anderberg (1973, S. 276ff.), Hartigan (1975, S. 215ff.), Späth (1975, S. 165ff.), Steinhausen/Langer (1977, S. 83ff.) Oder in Programmsysteme implementiert, z.B. BMDP (siehe Dixon, 1975, S. 307ff.) und Clustan (siehe Wishart, 1978, S. 31 ff.). Veröffentlichungen von Programmen zu den hierarchisch subdivisiven Verfahren finden sich nur vereinzelt, z.B. Späth (1975, S. 148 ff.), so daß man (vorerst) ersatzweise auf die unökonomische wiederholte Anwendung von Mengenerlegungsprozeduren zurückgreifen muß (Hinweise zu Programmen, siehe Abschnitt 3.2). - Eine Diskussion und partielle Evaluation zu Computerprogrammen für hierarchische Clusteranalysen findet man bei Aldenderfer/Blashfield (1978). Ergänzend sei hier auf einen beachtenswerten Gesichtspunkt der Programmanwendung hingewiesen (der auch in der methodologischen Literatur zumeist übergangen wird): Generell können während der Durchführung von (heuristischen) Verfahren Eindeutigkeitsprobleme auftreten. Bei agglomerativen Verfahren z.B. können in einer Matrix von Unähnlichkeitsmaßen zwischen Gruppen zwei (oder mehrere) minimale Eintragungen vorliegen, mit  $d(G_p, G_q) = d(G_p, G_r)$ . Es ist nun mit Hilfe eines begründeten Entscheidungskriteriums die Frage zu beantworten, ob die Gruppe p mit der Gruppe q oder mit der Gruppe r fusioniert werden soll. Wird dieses Problem in einem Programm nicht explizit thematisiert, so nimmt dieses bei der Ausführung zumeist automatisch eine Entscheidung vor, die vom Benutzer unbemerkt bleibt. Das Verfahren erfüllt dann ein relevantes methodisches Kriterium nicht, nämlich das der Unabhängigkeit des Resultats von der Reihenfolge der Objekte (zu diesem und weiteren Kriterien siehe Wright, 1974). Ein positives Gegenbeispiel aber geben z.B. Steinhausen/Langer (1977, S. 83), die einen programmierten Fehlerindikator einführen.

In der Forschungspraxis stellen sich im Zusammenhang mit der Anwendung von hierarchischen Clusteranalysen in verschiedenen Phasen des **Vorgehens** eine Reihe von Problemen, deren Darstellung im folgenden mit entsprechenden Lösungsansätzen versehen wird:

Für vorliegende Daten stellt sich vor der Anwendung von HCA-Verfahren die Frage nach deren Angemessenheit für die Repräsentation der Proximities. Damit wird für eine **a-priori-Evaluation** nach einem Entscheidungskriterium gesucht, welches die ‚clusteriness‘ (siehe Sneath 1969, S. 263) der Daten beurteilt. - Für den Fall gegebener Datenmatrizen (Beobachtungswiederholungen  $x$  (numerische) Merkmale) haben Oldenbürger/Becker (1976, besser verfügbar in Oldenbürger/Schwibbe, 1980) eine Strategie entwickelt, die eine statistische Prüfung der Clusterbarkeitshypothese gestattet. (Zu einem ähnlichen Vorgehen führt ein unabhängiger Vorschlag Sodeurs (1976) im Anschluß an Vogel (1975).) Dazu wurde eine Klasse von Maßen definiert, die die Abweichung der Proximities, z.B. Distanzen, von der ultrametrischen Ungleichung quantifizieren. (Passager wird ein einfacheres Maß auch z.B. von Carroll/Pruzansky 1980, S. 113 erwähnt.) Die Konstruktion des statistischen Tests erfolgte unter Verwendung der ‚Philosophie‘ der Randomization-Tests (Edgington 1969, 1980); dazu wurde eine geeignete Randomisierungsbasis gewählt, in der die Verteilungen der Variablen und damit ihre Kenngrößen erhalten bleiben. Für die statistische Entscheidung wird dann die Abweichung von der ultrametrischen Ungleichung für die empirischen Distanzen (hinsichtlich einer gewählten Modalität) mit der Verteilung der Abweichungen von der ultrametrischen Ungleichung für Distanzen aus zufälligen Datenmatrizen verglichen. (Diese werden durch zufällige Permutation der ursprünglichen Daten innerhalb der Variablen über die Beobachtungswiederholungen generiert.) - Für den Fall direkt erhobener Proximitymatrizen schlagen Sattath/Tversky (1977) die Untersuchung der Verteilung der Proximities vor. Sie konnten nämlich zeigen (Skewness-Theorem, 1977, S. 342ff.), daß die Verteilungen von Distanzen zwischen Punkten, die in euklidischen Räumen liegen, eher rechtsschief bzw. linkssteil sind, während Verteilungen von Distanzen zwischen Punkten in additiven Bäumen, z.B. Dendrogrammen, linksschief bzw. rechtssteil sind. (Dieser Ansatz ist allerdings bisher nicht mit einem statistischen Prüfverfahren gekoppelt, was aber durch Konstruktion eines Randomization-Tests für bestimmte Fälle relativ leicht geleistet werden kann. Dabei sind dann die tripelweisen Einschränkungen der Distanzen über die Dreiecksungleichung zu berücksichtigen.)

Hat man sich für die Durchführung der hierarchischen Clusteranalyse entschieden, so gibt es a-priori keine spezielle Methode der Wahl, deren Resultat eine optimale Repräsentation der Proximities garantiert. Es sollten also vielfältige, möglichst auch subdivisive Verfahren eingesetzt werden. Nach deren Anwendung stellt sich dann die Frage nach einer quantitativen **a-posteriori-Evaluation** der Resultate und damit einer nachvollziehbaren Auswahl der besten Repräsentation(en). Es gilt also, geeignete Goodness-of-Fit Kriterien für hierarchische Gruppierungen zu definieren (siehe z.B. Gower, 1975; eine Liste gibt Cormack, 1971, S. 338):

- Soll den Proximitydaten nur der Charakter von **Ranginformationen** unterstellt werden, so bietet sich als globales Maß für die Güte der Anpassung der ursprünglichen Proximities durch die aus den Dendrogramm rekonstruierten ultrametrischen Ähnlichkeits- bzw. Unähnlichkeitsmaße Goodman-Kruskals  $\gamma$  an (ein Rechenbeispiel gibt Lienert, 1978, S. 554ff.). Selbstverständlich dürfen auf diesen Index die üblichen statistischen Prüfverfahren nicht mehr angewendet werden. (Für kleine  $n (< 16)$  und die speziellen Verfahren des agglomerativen ‚Single-‘ und ‚Complete-Linkage‘ gibt Hubert (1974) Tabellen, die eine zufallskritische Beurteilung von  $\gamma$  gestatten. Wegen der Verfahrensabhängigkeit dieses Vorgehens ist die statistische a-priori-Prüfung der globalen Clusterbarkeitshypothese durch (approximative) Randomization-Tests wohl vorzuziehen.)

- Sind die Proximitymaße **metrisch** konstruiert worden oder wird ihnen metrischer Charakter zugeeignet, so bedarf es eines Zwischenschrittes vor der Anwendung globaler Goodness-of-Fit Kriterien. Wählt man z.B. die (kopphenetische) Korrelation  $r_{(emp d_{kl}, ult d_{kl})}$  (Hubert/Baker, 1977b), das ist die Produkt-Moment-Korrelation zwischen den Eintragungen der Dreiecksmatrizen (ohne Diagonale), oder die  $L_p$ -Abstandsnorm  $\sum_{k < l} | emp d_{kl} - ult d_{kl} |^p$ , mit  $p=2$  als

globales Maß für die Abbildungstreue der ursprünglichen Daten durch die konstruierte Ultrametrik, so ist dieser Index nur beziogl. des ‚Group-Average‘ oder seines divisiven Pendant als fair anzusehen (Farris, 1969). Alle anderen Verfahren generieren nämlich global und/oder partiell gedehnte oder gestauchte Dendrogramme bzw. Ultrametrien, so daß ein Vergleich der einzelnen Strukturierungsvorschläge (Dendrogramme) erst nach Beseitigung dieser Verfahrenseinflüsse durch Berechnung einer optimal angepaßten Ultrametrik  $ult d_{kl}$  erfolgen sollte. Dazu stehen die Techniken der numerischen Approximation gegebener Baumstrukturen an Proximitymatrizen zur Verfügung, wie sie etwa bei Eisler (1973) oder Roskam (1975, S. 49f.) für die generelle Anwendung beschrieben werden. Im hier zur Debatte stehenden Fall aber reicht es aus, die  $(L_2)$ -Knotenhöhe jeweils durch Mittelung der empirischen Proximities anzugeben, deren zugehörige Objektpaare durch den Knoten erst zusammengeführt werden. (Bei Wahl der robusteren  $L_p$ -Abstandsnorm mit  $p = 1$  wäre hier die Angabe des Medians angemessen.) Nach dieser numerischen Anpassung können die strukturellen Verfahrensvorschläge (Dendrogramme) verbessert graphisch dargestellt und ihre Abbildungsleistung durch Berechnung von z.B.  $r_{(emp d_{kl}, ult d'_{kl})}$  quantitativ bewertet werden.

Hat man nach Berechnung der Globalmaße eine begründete Auswahl unter den Repräsentationen getroffen, so verbleibt als Aufgabe die spezifischere Untersuchung der Abbildungsleistung des HCA-Resultats. Diese kann durch Inspektion der Rangplatzdifferenzen oder der Residuen  $emp d_{kl} - ult d'_{kl}$  erfolgen; das Vorzeichen solcher Werte gibt dann an, ob die jeweiligen Objektpaare

durch die clusteranalytische Darstellung als zu stark separiert oder als zu eng zusammengehörig ausgewiesen werden; die absolute Größe solcher Abweichungswerte zeigt auf, an welchen Stellen die kompakte Repräsentation durch hierarchische Gruppierung erhebliche ‚Fehlleistungen‘ zeitigt. - Eine solche (exploratorische) Bemühung kann einerseits undifferenzierte Globalinterpretationen von Dendrogrammen vermeiden helfen, andererseits könnte die Verteilung (das ‚Pattern‘) der Abweichungen Hinweise auf (weitere) alternative Repräsentationsmodi der Proximitydaten liefern.

Sollen in der Anwendung verschiedene hierarchische Strukturierungen verglichen werden, so ist dazu auf folgende Ansätze hinzuweisen:

- Boorman/Olivier (1973) definieren eine Reihe von Metriken über Folgen von Partitionen bzw. über Ultrametrien, die z.B. zur Auswahl eines ‚mittleren‘, maximal konsistenten Dendrogramms herangezogen werden sollen.
- Zur konfirmatorischen statistischen Prüfung der Übereinstimmung a-priori formulierter hierarchischer Strukturhypothesen mit den Resultaten der Anwendung von HCA-Verfahren auf (unabhängige) Daten setzen Hubert/Baker (1977) das äußerst flexible Paradigma des ‚quadratischen Assignierens‘ ein (Lawler, 1975; Hubert/Schultz, 1976). Diese generelle Strategie der Datenanalyse ist über die Untersuchung von Permutationsverteilungen eng mit der Methodologie der Randomization-Tests verknüpft. Sie gestattet auch eine konfirmatorisch zufallskritische Beurteilung der Übereinstimmung von (z.B. nach Verfahrensanwendung ultrametrischen) Proximitymatrizen aus verschiedenen Untersuchungsgruppen (Subsamples). Gleichzeitig liefert die Theorie des quadratischen Assignierens eine Grundlage zur probabilistischen Evaluation der von Boorman/Olivier (1973) vorgeschlagenen Distanzmaße für Folgen von Partitionen. - Bei der Anwendung dieser komfortablen Evaluations- und Prüfstrategie ist zur Vermeidung von Artefaktinterpretationen grundsätzlich zu beachten, daß die zu vergleichenden Proximitymatrizen unabhängiger Herkunft sind. (So wäre z.B. die ‚Prüfung‘ der kophenetischen Korrelation zwischen empirischer Proximitymatrix und der nach Verfahrensanwendung rekonstruierten Ultrametrik unzulässig.) - Erhebliche Weiterentwicklungen des Paradigmas des quadratischen Assignierens durch die Einführung und Untersuchung des Vergleichs genereller Proximityfunktionen werden von Hubert/Baker (1977), Hubert (1978) und Hubert/Baker (1979) angeregt. Im hier gegebenen Diskussionszusammenhang ist die Möglichkeit des simultanen Vergleichs  $K$  unabhängiger Proximitymatrizen von besonderem Interesse (Hubert, 1979c; siehe auch Hubert, 1979a, 1979b). An dieser Stelle muß aber ebenso auf die Ansätze zur Prüfung von Strukturhypothesen zu Korrelations- und Kovarianzmatrizen in den Arbeiten Jöreskogs (z.B. 1978) hingewiesen werden (siehe auch Steiger, 1980).

Abschließend sei hervorgehoben, daß die Ausführungen in diesem Kapitel sich vornehmlich auf die Anwendung von HCA-Verfahren zur Konstruktion von Folgen von Partitionen mit Gruppen lagehomogener Objekte bezogen. Diese Einschränkung aber ist weder notwendig noch sinnvoll. Nach Definition geeigneter Proximitymaße und/oder Optimierungskriterien können Verfahren, die hierarchische Repräsentationen generieren, für **verschiedenste Zielsetzungen** verwendet werden. Als Beispiele für weitere Entwicklungen seien die folgenden Ansätze genannt:

- Bei Vorliegen nominaler Datenmatrizen kann die hierarchische Entropieanalyse der ‚Canberra-Schule‘ eingesetzt werden (z.B. Williams/Lance, 1977).
- Wendet man die Hierarchisierung auf allgemeinere Problemstellungen der Mengenerlegung an, wie sie zum Abschluß des Abschnitts 3.2 aufgewiesen werden (z.B. Aufteilung nach Distanz zu Hyperebenen, ‚Klassifikationsansatz‘), so erhält man weitere Verfahrensklassen, für die in erster Linie subdivisives Vorgehen adäquat sein wird.
- Die Aufhebung der Einschränkung, hierarchische Clusteranalysen auf symmetrische Proximitymatrizen anzuwenden (ohne diese einfach durch Mittelung korrespondierender Eintragungen zu symmetrisieren,) nimmt Hubert (1973) vor (siehe auch Hubert, 1976 und Baker/Hubert, 1977).
- In Abhebung von der sonst üblichen ‚ungerichteten‘ Clusteranalyse entwickelten Kleiter/Timmermann (1977) einen Algorithmus zur gerichteten hierarchischen Voraussetzungs-Struktur-Analyse. Eine ausführliche Darstellung dieser inhaltlich-theoretisch interessanten Repräsentationsverfahren am Beispiel der Analyse sachlogischer und empirischer Lernvoraussetzungen (Lernwege) geben Kleiter/Petermann (1977, mit Computerprogrammen, siehe auch Kleiter, 1980).

In den folgenden Abschnitten werden einige Ansätze dargestellt, die den Rahmen der Anwendung clusteranalytischer Verfahren auf die Untersuchung weiterer diskreter mathematischer Strukturen ausdehnen.

### 3.4 Baumrepräsentationen und hybride Modelle

Zusammenhängende, kreislose Graphen werden als Bäume bezeichnet (siehe z.B. Berge, 1976, S. 24f.; Harary, 1974, S. 42f.). Die Verbindung dieses Konzepts mit der Clusteranalyse ergibt sich insbesondere aus der Möglichkeit, die durch Anwendung von Verfahren der hierarchischen Clusteranalyse generierten Dendrogramme als spezielle Bäume anzusehen. Die daran anknüpfbaren Weiterentwicklungen der Baumrepräsentation von Proximitymatrizen (Patrinos/Hakimi, 1972) sollen im folgenden kurz angesprochen werden.

Faßt man Dendrogramme als ultrametrische Bäume auf (im Unterschied zur nicht-metrischen Interpretation als ‚bare trees‘, siehe Boorman/Olivier, 1973), so sind folgende Einschränkungen charakteristisch:

- (a) Alle Objekte erscheinen als Endpunkte bzw. terminale Knoten des Baums und
- (b) die Entfernungen zwischen allen Objekten aus elementfremden Gruppierungen sind gleich groß.

Lockert man die Restriktionen und verlangt lediglich, daß den Kanten des Baumes positive Gewichtszahlen zugeordnet werden, deren Addition die Länge des Weges (Kantenfolge) zwischen den jeweiligen Objektpaaren angibt, so entstehen die sogenannten freien oder additiven Bäume. Die durch sie abgebildeten Distanzen folgen der sogenannten Vierpunktmetrik (Bunemann, 1971):

Für alle  $o_k, o_l, o_r, o_s \in 0$  gilt:

$$(D. 6) \quad d(o_k, o_l) + d(o_r, o_s) \leq \max \{d(o_k, o_r) + d(o_l, o_s), d(o_k, o_s) + d(o_l, o_r)\}$$

(vgl. Abschnitt 2; es gilt ferner: (D. 5)  $\Rightarrow$  (D. 6)  $\Rightarrow$  (D. 4)). Algorithmen zur Repräsentation von Proximitymatrizen durch additive Bäume sowie Anwendungsbeispiele geben Cunningham (1974, 1978), Sattath/Tversky (1977), Carroll/Chang (1973) und Carroll (1967) sowie Carroll/Pruzansky (1980). In engstem Zusammenhang mit diesem strukturellen Ansatz stehen die fundamentalen theoretischen Diskussionen des Feature-Modells der Ähnlichkeit durch Tversky (1977) und des probabilistischen Choice-Modells der Präferenz durch Tversky/Sattath (1979). Colonius/Schulze (1977, 1979) geben notwendige und hinreichende Bedingungen für die Baumrealisierbarkeit einer empirischen, nicht-numerischen Struktur an und schlagen eine Konstruktionsmethode vor. - Boyd/Wexler (1973) entwickeln eine Komponentenanalyse strukturierter Bäume (mit ungerichteten Kanten) und wenden sie auf die semantische Analyse von Verben an. - Läßt man für additive Bäume die Einschränkung fallen, daß die positiv gerichteten Kanten ungerichtet sind und ersetzt diese durch zwei gerichtete Kanten mit verschiedenen Gewichten, so erhält man sogenannte gewichtete bidirektionale Bäume. Cunningham (1978) benennt notwendige und hinreichende Bedingungen für deren Existenz und Eindeutigkeit und zeigt die Möglichkeiten der Repräsentation unsymmetrischer Proximitymatrizen durch bidirektionale Bäume an einem Beispiel aus dem Bereich des ‚sentence memory‘ auf.

Erste Vorschläge zur Repräsentation von Proximitydaten durch **gemischte bzw. hybride Modelle** gehen auf Degerman (1970, 1972) zurück. Diese kombinieren klassifikatorische Darstellungen mit denen der mehrdimensionalen Skalierung. (So unterscheiden sich beispielsweise Tiere einerseits hinsichtlich diskreter Charakteristika (z.B. ‚ist Säugetier‘ oder ‚ist nicht Säugetier‘), andererseits hinsichtlich kontinuierlicher Merkmale (z.B. Größe).) In der erweiterten

Fassung von Carroll (1976) werden gegebene Proximitymatrizen, z.B. von Unähnlichkeiten  $D_{\text{emp}}$ , als additive Kombinationen verschiedener (theoretischer) Proximitymatrizen dargestellt:

$$\text{HM:} \quad D_{\text{emp}} \simeq D_1 + D_2 + D_3 + \dots D_m$$

(, $\simeq$ ' steht für ‚approximiert nach einem Optimalitätskriterium‘). Dabei stehen die  $D_1$  bis  $D_m$  für Proximitymatrizen, die z.B. Abstände aus (mehr)dimensionaler Skalierung, ultrametrischen Bäumen und/oder additiven Bäumen darstellen. Den überzeugenden Erfolgen dieses generellen Ansatzes bei der Repräsentation exemplarischer Daten (siehe auch Carroll/Pruzansky, 1980) und der Verfügbarkeit algorithmischer Lösungen, insbesondere durch alternierende Kleinst-Quadrat-Approximationen (siehe dazu auch die Referenzen in Young/de Leeuw/Takane, 1980), stehen z.Z. noch Probleme gegenüber, die auf die immense Flexibilität der Repräsentation und die entsprechend große Zahl der zu schätzenden Parameter zurückgehen; sie betreffen vor allem die Eindeutigkeit und Identifizierung der Modelle sowie deren Prüfbarkeit.

### 3.5 Überlappende Gruppierung

Die **bisher** in diesem Kapitel behandelten Ansätze zur Erzeugung homogener Gruppen verlangen, daß diese paarweise elementenfremd bzw. disjunkt sind:

$$\text{Für alle } p, q \text{ gilt: } G_p \cap G_q = \emptyset.$$

Zwar sind in der Anwendung deutlich überwiegend Verfahren eingesetzt worden, die diese Einschränkung vornehmen, es lassen sich aber leicht Fälle angeben, die eine Einteilung von Objekten in einander ausschließende Klassen unangemessen erscheinen lassen. (So stellt z.B. die Liste wissenschaftlicher Publikationen unter sogenannten ‚key-words‘ im Index der ‚Psychological Abstracts‘ eine nichtdisjunkte Gruppierung dar; es wäre nicht praktisch, würde man jede Arbeit nur unter genau einem Schlagwort finden können (Zerlegung).) In der methodologischen Literatur findet sich eine Fülle von Verfahren, die für eine Objektmenge mit gegebener (symmetrischer) Proximitymatrix zumindest eine Menge  $A = \{G_1, G_2, \dots, G_p, G_q, \dots\}$  generieren, deren Elemente nichtdisjunkte Gruppierungen sind. Da diese Konstruktionsmethoden in der Regel ohne Optimalitätskriterium arbeiten, müssen sie als heuristisch bezeichnet werden. Neben dem notwendig hohen Aufwand liegt wohl auch in dieser Eigenschaft ein Grund, warum sie in der Anwendung bisher so wenig Beachtung fanden: Es liegt im allgemeinen kein Maßstab vor, nach dem alternative Klassifikationssysteme  $A_i$  als mehr oder weniger angemessen zur Repräsentation einer Objektmenge und deren Proximities beurteilt werden können. Von gegenwärtigen Verfahrensentwicklungen aber wird dieser Gesichtspunkt stärker berücksichtigt.

Die wohl ausführlichste Übersicht der theoretischen Grundlagen und der vielfältigen Verfahren zur überlappenden Gruppierung gibt Bock (1974, S. 316ff.). - In diesem Abschnitt wird dagegen die kurze Diskussion auf drei zentrale Ansätze eingeschränkt: die Untersuchung maximaler Cliques, die Generierung überlappender Cluster nach Jardine/Sibson und das ADCLUS-Modell von Shepard/Arabie.

Für die Betrachtung der Aufgabenstellung unter dem Stichwort **„maximale Cliques“** sei zu einer Objektmenge eine Matrix von Unähnlichkeiten  $[d_{kl}]$  vorgegeben. Unter einer Clique der Stufe  $c$  (cut) versteht man eine Menge  $G_p \subset \mathbf{0}$ , deren sämtliche Paare von Elementen einander nicht unähnlicher sind als durch  $c$  angegeben wurde:

Für alle  $o_k, o_l \in G_p$  gilt  $d_{kl} \leq c$ .

„Maximal“ heißt die Clique  $G_p$ , wenn es in der Restmenge  $\mathbf{0} \setminus G_p$  kein weiteres Objekt  $o_r$  gibt, dessen Unähnlichkeiten zu allen Objekten in der Clique kleiner/gleich  $c$  sind. Auf beinahe natürliche Weise führen diese Begriffsbestimmungen zu einem - auf den ersten Blick - einfachen heuristischen Konstruktionsverfahren überlappender Gruppierungen: (1) Man legt einen geeigneten Wert  $c$  fest und (2) sucht alle maximalen Cliques der Stufe  $c$ . Für beide Schritte gibt es algorithmische Lösungsansätze:

- (1) Faßt man beispielsweise die Unähnlichkeiten als metrische Größen auf, so kann ein optimaler ‚cut‘ z.B. mittels des Fisher-Algorithmus (siehe Abschnitt 3.2) erfolgen.
- (2) Für das nicht-triviale Problem der Feststellung aller maximalen Cliques der Stufe  $c$  steht dann z.B. der effektive ‚branch-and-bound‘-Algorithmus von Bron/Kerbosch (1973, weitere Angaben bei Shepard/Arabie, 1979, S. 99) zur Verfügung.

Allerdings ist dieses praktikable Vorgehen wohl als recht vordergründig anzusehen, da zu den genannten Schritten allgemeinere Probleme korrespondieren, die bei der Herstellung von Verfahren der nichtdisjunktiven Clusteranalyse zentral sind:

- (1) Wieviele Werte  $c_t$  sollen verwendet werden, um möglicherweise quasihierarchische Varianten des Verfahrens zu erzeugen?
- (2) Warum soll man sich auf den Gruppierungsbegriff der ‚maximalen Clique‘ festlegen? (Dieser garantiert zwar, daß der Durchmesser einer Gruppierung höchstens  $c_t$  ist (wie bei HCA-Verfahren das ‚Complete Linkage‘), man könnte aber ebenso begründet Gruppen generieren, deren Elemente zu allen anderen Objekten der Gruppe z.B. höchstens eine durchschnittliche Unähnlichkeit von  $c_t$  aufweisen.)

Bei Betrachtung der ersten Fragestellung ist folgende Argumentation relevant: Zu einer gegebenen Matrix von Unähnlichkeiten können sinnvoll höchstens  $\binom{n}{2}$  Cutwerte  $c_i$  untersucht werden. Geht man diese der Reihe nach durch und generiert die (z.B.) jeweils neu entstehenden maximalen Cliques, so stellt dieses Mengensystem eine Repräsentation der Proximitymatrix ohne Informationsverdichtung dar, also eine reine Datendeskription. (Die Begründung der Gültigkeit dieser Aussage kann parallel zum Beweis des wichtigen Dekompositionstheorems der ‚Fuzzy Relations‘, siehe Kaufman (1975, S. 67ff. und s. 387ff.), erfolgen.) Eine strukturierte quasi-hierarchische Repräsentation von Proximities entsteht deshalb erst durch Einführung **spezieller Restriktionen**. Da es dafür verschiedene Möglichkeiten gibt, schlagen u.a. Jardine/Sibson (1968, 1971, S. 59ff., siehe auch Bock, 1974, S. 423ff.) drei Verfahrensklassen vor, deren bekannteste das sogenannte  $B_k$ -Clustering ist (siehe auch Cole/Wishart, 1970; Rohlf, 1974, 1975; Day, 1977). Für diese Methoden ist als Einschränkung die erlaubte Anzahl von Elementen in den Überlappungen  $(k-1)$  charakteristisch. Wegen ihrer engen Beziehung zu dem von Jardine und Sibson (1971) favorisierten hierarchischen ‚Single Linkage‘ Verfahren setzen sich diese Methoden allerdings der Kritik aus, Verkettungseffekten zu unterliegen. Listen von Fortran-Programmen geben Jardine/Sibson (1971, S. 227ff., u.a. auch zu Sibsons C-Verfahren, welche die erlaubte Größe des Durchmessers der Überlappung einschränken); die  $B_k$ -Methoden sind auch in Wisharts Programmpaket Clustan, 1978, S. 65ff., enthalten. Anwendungen dieser Verfahren sind bisher kaum bekannt (siehe aber Friendly, 1977).

Ein der zweiten Fragestellung zugeordneter Ansatz zur Generierung überlappender, nicht notwendig hierarchischer Gruppierungen stammt von Shepard/Arabe (1979). Das ‚additive clustering‘ geht von Ähnlichkeitsmaßen in der Proximitymatrix aus und repräsentiert diese durch ein einfaches Modell:

$$\hat{s}_{kl} = \sum_j w_j p_{kj} p_{lj}$$

Darin sind die  $w_j$  positive Gewichte, die die Sättigung einer Eigenschaft  $j$  angeben, während die  $p_{kj}$ ,  $p_{lj}$  zeigen, ob z.B. ein Objekt  $k$  die Eigenschaft  $j$  besitzt ( $p_{kj}=1$ ) oder nicht ( $p_{kj}=0$ ). Die Ähnlichkeiten zwischen den Objekt-paaren werden also als Kombinationen der Gewichte diskreter Objektmerkmale dargestellt. Einen Algorithmus zur Lösung dieses Dekompositionsproblems unter Verwendung eines Maßes der erklärten Varianz als Optimierungskriterium entwickelten Shepard/Arabe (1979) unter dem Namen ADCLUS. Eine effizientere Alternative beschreiben Arabe/Carroll (1980). Sie verwenden Techniken der ‚mathematischen Programmierung‘ zur Lösung des Optimierungsproblems: MAPCLUS. Dabei scheinen sie im allgemeinen zu sparsameren Lösungen zu gelangen, d.h. zur Repräsentation der Ähnlichkeiten werden weniger Attribute benötigt. In den zitierten Arbeiten werden zahlreiche Beispiele für Ergebnisse der Verfahrensanwendung auf empirische Daten gege-

ben. Schließlich erweitern Arabie/Carroll (1980) das ADCLUS-Modell auf die Möglichkeit der Analyse individueller Differenzen (INDCLUS). - Geeignete statistische Prüfverfahren für diese Ansätze sind noch zu entwickeln. Die Residueninspektion im Anwendungsfall ist selbstverständlich.

### 3.6 Cluster in Datenmatrizen

Aufgabenstellung dieser Verfahren ist das ‚Auffinden‘ von Blöcken, Feldern, ‚Pattern‘ oder Strukturen gleicher oder ‚ähnlicher‘ Daten in  $n \times m$  - Matrizen  $[x_{ij}]$  (Beobachtungswiederholungen  $\times$  Merkmale) durch geeignete Umordnung ihrer Zeilen und/oder Spalten. Während die in den bisherigen Abschnitten betrachteten Vorgehensweisen Gruppen ähnlicher Objekte aus **einer** Grundmenge konstruieren, untersuchen clusteranalytische Verfahren zur Strukturierung von Datenmatrizen Mengen von Objekten hinsichtlich **zweier** Modalitäten (simultan).

Kritisch für die Wahl bzw. Einsatzmöglichkeit solcher Verfahren ist die Beantwortung der Frage nach der Vergleichbarkeit kategorialer und/oder numerischer Merkmalsausprägungen (Hartigan, 1972, S. 124). Der genuine Anwendungsfall der Clusteranalyse von Datenmatrizen ist gegeben, wenn Vergleichbarkeit der Ausprägungen hinsichtlich der Merkmale **und** der Beobachtungswiederholungen vorliegt, bzw. durch geeignete Transformation (z.B. Standardisierung) hergestellt werden kann (Beispiele: (a) Matrix Bundesländer  $\times$  Geburtsjahrgänge mit Eintragungen, die den prozentualen Anteil der Personen mit Abitur angeben, (b) Personen  $\times$  Items eines Tests mit den Ausprägungen ‚gelöst‘ (= 1) und ‚nicht gelöst‘ (=0)).

Sind die Ausprägungen hinsichtlich einer Modalität (z.B. zwischen den Klassen verschiedener nominaler Merkmale) inkompatibel, so ist zumeist auf die in Abschnitt 3.1 bis 3.4 beschriebenen clusteranalytischen Verfahren zurückzugreifen (siehe ergänzend z.B. Hartigan, 1975, S. 257, aber auch Generalisierungsversuche ebd., S. 288ff. und deren Weiterentwicklung S. 299ff.).

Liegen bimodal vergleichbare Daten vor, so kann die Problemstellung der Verfahren der sogenannten direkten Clusteranalyse (Hartigan) oder der Konstruktion von Blockmodellen (Arabie/Boorman/Levitt, 1978; White, 1977) wie folgt formuliert werden: Gesucht ist sowohl eine Permutation der Zeilen- als auch der Spaltenindizes, so daß die (umgestellte) Datenmatrix bezüglich (der Einfachheit) ihrer inneren Strukturierung optimal ist. Beispiele für Zielfunktionen sind: (a) die Anzahl identischer Ausprägungen benachbarter Eintragungen bei kategorialen Merkmalen oder (b) die Summe der absoluten Differenzen benachbarter Eintragungen bei numerischen Merkmalen. (Zu beachten ist hier die (in der Literatur vernachlässigte) Möglichkeit der Einführung von ‚Fen-

stern' verschiedener Größe.) - Essentieller Bestandteil solcher Optimierungsverfahren ist die Konstruktion von Reihenfolgen (Seriation), deren algorithmische Untersuchung als Standardproblem des 'Operations Research' wohlbekannt ist (Müller/Merbach, 1970; Christofides, 1975, 1975b, S. 236ff.) und erst durch Hubert (1976), Baker/Hubert (1977) und Hubert/Baker (1978) in die methodologische Diskussion der Psychologie eingebracht wurde.

Beispiele für die Verfahrensanwendungen der direkten Clusteranalyse bzw. der Blockmodellierung für empirische Daten geben Hartigan (1975, S. 262 und S. 283 f.) und Arabie/Boorman/Levitt (1978).

Programmsprachliche Formulierungen von Verfahren (ohne besondere Hervorhebung des generellen Seriationsansatzes) teilt Hartigan (1975, S. 271 ff. und S. 294ff.) mit. Von diesem Autor stammt auch ein Programm, welches in das BMDP-Paket (Dixon, 1975, S. 339ff.) unter dem Namen 'Block-Clustering' implementiert wurde. (Dabei handelt es sich wohl um eine ältere Version; es ist aber auf die instruktive Erklärung des Verfahrens (S. 352f.) und den 'Sample-output' (S. 354f.) hinzuweisen.)

Die Verfahren der direkten Clusteranalyse von Datenmatrizen bzw. des 'Block-Clustering' stellen insbesondere für exploratorische Zwecke hochinteressante Instrumente dar: sie strukturieren die ursprünglichen Daten und stützen die Anschauung. Da darin auch Gefahren liegen, ist die Wichtigkeit probabilistischer und statistischer Evaluationen besonders hervorzuheben (Ansätze geben z.B. Hartigan, 1975, S. 285 und White, 1977). Abweichungen der Daten von 'Modellprädiktionen durch die Blöcke' sind in den Resultaten der Verfahrensanwendung ('Print-outs') direkt ersichtlich.

#### *4. Diskussion und Ausblick*

Obwohl sich die vorliegende Arbeit in der Darstellung auf die Hauptgesichtspunkte clusteranalytischer Verfahren einzuschränken hatte, ist wohl auch hier deutlich geworden: Es gibt eine immense Vielfalt von Zielsetzungen, Problemstellungen, Ansätzen, Methoden und Verfahrensvarianten zur Clusteranalyse. - Dem steht in der Forschungspraxis z.Z. ein erheblich eingeschränkter Gebrauch von hauptsächlich ca. drei Verfahren der agglomerativen hierarchischen Clusteranalyse gegenüber. - Will man diese Diskrepanz erklären, so wäre der Verweis auf die generelle Angemessenheit der HCA-Verfahren für die Klärung verschiedenster inhaltlich-theoretischer Fragestellungen sicher nicht akzeptabel. Daneben scheinen auch Argumente, die die ökonomische Durchführbarkeit dieser Methoden und die mangelnde Verfügbarkeit von Computerprogrammen zu alternativen Verfahren benennen, allein nicht überzeugen zu können.

Vermutlich kommen zumindest zwei Merkmale der Diskussionsentwicklung in der Literatur hinzu: (a) Nach dem Aufweis des methodischen Paradigmas durch den Artikel von Johnson (1967) sind die agglomerativen HCA-Verfahren als ‚Allroundinstrumente‘ (ähnlich dem ‚Little Jiffy‘ in der Faktorenanalyse, d.i. Hauptkomponentenextraktion und Varimaxrotation) üblich geworden. (b) Die ‚Landschaft‘ alternativer Ansätze und Verfahren in der methodologischen Literatur ist wegen ihres Umfangs und Aspektreichtums zu komplex und damit unüberschaubar geworden. Sie bietet kein geordnetes Bild, sondern macht eher einen diffusen Eindruck, der dem interessierten Anwender einen zielgerichteten Zugriff nicht erlaubt.

Ein Beleg für die Unübersichtlichkeit der mathematischen Darstellungen ist das ‚Sprachengewirr‘ der Formalisierungen. So findet man z.B. verbandstheoretisch (Leuschner, 1973), mengentheoretisch (Hubert, 1977a), matrixtheoretisch (Kim/Roush, 1978), graphentheoretisch (Hubert, 1974, 1977b; Matula, 1977) und schließlich ‚Fuzzy Set‘-theoretisch (Negoița/Ralescu, 1975, S. 169ff.; Yeh/Bang, 1975; Zadeh, 1977) geprägte Präsentationen der Clusteranalyse. - Ein weiteres Merkmal der methodologischen Literatur, welches die Orientierung erschwert, ist die vorwiegende **Betonung der rein verfahrenstechnischen Aspekte** des Prozesses und der Ergebnisse von Clusteranalysen. (Es wird gezeigt, was man mit Daten (so alles) tun kann, und was dabei herauskommt.) Dabei wird der für den Einsatz in einer empirischen Wissenschaft **zentrale Modellaspekt vernachlässigt**. Dieser betont, daß Ergebnisse von Repräsentationsverfahren Daten abzubilden haben, und nicht umgekehrt die Daten lediglich Ausgangspunkt von Verfahrensanwendungen sind.

Aus dem Interesse, den Modellaspekt der Resultate von Verfahrensanwendungen hervorzuheben, verstehe ich unter ‚Clusteranalyse‘ Methoden, die die Approximation einer Teilklasse (numerisch evaluierter) diskreter mathematischer Modelle an empirische Daten leisten.

Diese Sichtweise führt für die Entwicklung der methodologischen Literatur zu folgenden Forderungen:

- Da bei der Approximation von Modellen grundsätzlich deren Angemessenheit in Frage steht, sollten verstärkt vor allem **überprüfungsverfahren** entwickelt und dem potentiellen Benutzer auch in praktikabler Form zur Verfügung gestellt werden (Veröffentlichung/Austausch von Programmen; Implementation zumindest in die weit verbreiteten Pakete). Es ist nicht weiter akzeptabel, daß Programme zur Durchführung clusteranalytischer Verfahren gewöhnlich keinerlei Indizes zur Beurteilung der statistischen und/oder praktischen Signifikanz bzw. Bedeutsamkeit des Resultats mitteilen und eine Untersuchung von Residuen nicht vorgesehen ist (z.B. auch bei Programmen zur Faktoranalyse).
- Die Approximation der diskreten Modelle sollte grundsätzlich als Problem der **Optimierung einer Zielfunktion** (unter Nebenbedingungen) angesehen

werden. (Dies könnte das ‚Gewirr‘ mathematischer Sprachen und die ‚Anhäufung‘ heuristischer Methoden abbauen helfen.) Solche Optimierungsprobleme werden bereits seit einiger Zeit in der mathematischen Verfahrensforschung (Operations Research) unter den Überschriften ‚Integer-Programmierung‘ und ‚diskrete‘, ‚kombinatorische‘, sowie ‚mathematische Programmierung‘ diskutiert. (Deshalb sei ausdrücklich auf die Bibliographien von Kasting (1976) und Hausmann (1978) und stellvertretend auf die Periodica ‚Computing‘ und ‚Mathematical Programming‘ hingewiesen.) Zwar gibt es einige Autoren, die diesen Ansatz auch im Bereich der Clusteranalyse verfolgt haben (z.B. Ward, 1963; Hartigan, 1967; Rao, 1971; Brucker, 1978; sowie Carroll und Hubert in verschiedenen Arbeiten), in der psychologischen mathematisch-statistischen Literatur hat sich aber die selbstverständliche Beachtung von Ansätzen zur Modellapproximation durch Optimierungsverfahren (siehe z.B. Horst, 1979; Dixon/Spedicato/Szegö, 1980) noch nicht allgemein durchgesetzt (zu ihrer Bedeutung in der Statistik siehe Kennedy/Gentle, 1980).

Das übliche **forschungspraktische Vorgehen** nach dem Schema: Datenerhebung/Datenerfassung/Programmanwendung (mit Standardoptionen)/Interpretation der Resultate sollte bei verstärkter Beachtung des Modellaspekts von Ergebnissen clusteranalytischer Repräsentationsverfahren durch einige Kontrollüberlegungen bereichert werden:

- Es ist die Frage zu beantworten, ob die Approximation eines diskreten Modells Gegenstand und Zielvorstellung der Untersuchung angemessen ist. (Für die HCA wird dieses Problem von Kleiter/Fillbrandt, 1973, behandelt.) Durch theoretisch inhaltliche Überlegungen ist also zu begründen, warum die Repräsentation von untersuchten Objekten, z.B. durch eine Mengenerlegung, adäquat sein soll. (Durch den Einsatz eines Verfahrens wird die Gruppierungsdarstellung erzwungen.)
- Die Geltung der Voraussetzungen von Verfahrensbestandteilen ist für vorliegende Daten zu prüfen (z.B. die Linearität der Zusammenhänge bei Verwendung von Korrelationen oder euklidischen Distanzen als Proximities).
- Die Gültigkeit der diskreten Modelle ist (soweit möglich) zumindest einer konfirmatorischen statistischen Prüfung zu unterziehen (s.O., z.B. Abschnitte **3.2** und 3.3).
- Liegen zur Approximation von diskreten Modellen lediglich heuristische Verfahren (z.B. HCA) vor, so sind stets mehrere Lösungen zur Repräsentation der Daten zu generieren und mittels eines quantitativen Maßes der Anpassungsgüte zu bewerten.
- Schließlich sind die Abweichungen zwischen den Daten und der Modelldarstellung im einzelnen zu untersuchen, denn - wie John W. Tukey sagt -:

„Residuals are our main tool in going further. They are to the data analyst what powerful magnifying glasses, sensitive chemical tests for bloodstains, and delicate listening devices are to a story-book detective. They permeate all sorts of data analysis and appear in many guises.“ (1977, S. 125)

## *Literatur*

Diese Liste enthält nur Titel, die in der vorliegenden Arbeit zitiert wurden. - Bibliographischen Charakter haben die Literaturverzeichnisse in Bock (1974), Sneath/Sokal (1973) und Vogel (1975). Für neuere Arbeiten siehe ergänzend Eckes/Rosbach (1980).

- Abadie, J. (Ed). 1970. Integer and nonlinear programming. Amsterdam: North-Holland.
- Ahrens, H. J. 1974. Multidimensionale Skalierung. Weinheim: Beltz.
- Aldenderfer, M. S. & Blashfield, R. K. 1978. Computer Programms for performing hierarchical cluster analysis. *Applied Psychological Measurement*, 2, 405-413.
- Anderberg, M. R. 1973. Cluster analysis for applications. New York: Academic Press.
- Arabie, Ph. & Boorman, S. A. 1973. Multidimensional scaling of measures of distance between partitions. *Journal of Mathematical Psychology*, 10, 148-203.
- Arabie, Ph., Boorman, S. A. & Levitt, P. R. 1978. Constructing blockmodels: How and why. *Journal of Mathematical Psychology*, 17, 21-63.
- Arabie, Ph. & Carroll, J. D. 1980. MAPCLUS: A mathematical programming approach to fitting the ADCLUS model. *Psychometrika*, 45, 211-235.
- Atchley, W. R. & Bryant, E. H. (Eds). 1975. Multivariate statistical methods. Among-groups covariation. Stroudsburg: Dowden, Hutchinson & Ross, Inc.
- Baker, F. B. 1974. Stability of two hierarchical grouping techniques. Case 1: Sensivity to data errors. *Journal of the American Statistical Association*, 69, 440-445.
- Baker, F. B. & Hubert, L. J. 1975. Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association*, 70, 31-38.
- Baker, F. B. & Hubert, L. J. 1976. A graph-theoretic approach to goodness-of-fit in complete-link hierarchical clustering. *Journal of the American Statistical Association*, 71, 870-878.
- Baker, F. B. & Hubert, L. J. 1977. Applications of combinatorial programming to data analyses: Seriation using asymmetric proximity measures. *British Journal of Mathematical and Statistical Psychology*, 30, 154-164.
- Balas, E. & Padberg, M. W. 1975. Set partitioning: A survey. In: Roy, B. (Ed.). 205-258.
- Banfield, C. F. & Basill, L. C. 1977. Algorithm AS 113: A transfer algorithm for non-hierarchical classification. *Applied Statistics*, 26, 206-210.
- Beale, E. M. L. 1970. Selecting an optimal subset. In: Abadie, J. (Ed.). 451-562.
- Beale, E. M. L., Kendall, M. G. & Mann, D. W. 1967. The discarding of variables in multivariate analysis. *Biometrika*, 54, 357-366.

- Berge, C. 1973, 1976<sup>2</sup>. Graphs and hypergraphs. Amsterdam: North-Holland.
- Blashfield, R. K. 1976. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, 83, 377-388.
- Blashfield, R. K. 1980. The growth of cluster analysis: Tryon, Ward, and Johnson. *Multivariate Behavioral Research*, 15, 439-458.
- Blashfield, R. K. & Aldenderfer, M. S. 1978. The literature on cluster analysis. *Multivariate Behavioral Research*, 13, 271-295.
- Blashfield, R. K. & Aldenderfer, M. S. 1978. Computer Programms for performing iterative partitioning cluster analysis. *Applied Psychological Measurement*, 2, 533-541.
- Block, H. G. & Böck, G. et al. 1974. Klassifizierungsprozesse bei der Auswertung von Röntgenaufnahmen. In: Klix, F., Sydow, H. & Wysotzki, F. (Eds). 157-178.
- Bock, H. H. 1974. Automatische Klassifikation. Göttingen: Vandenhoeck & Ruprecht.
- Bonner, R. E. 1964. On some clustering techniques. *IBM Journal of Research and Development*, 8, 22-32.
- Boorman, S. A. & Arabie, Ph. 1972. Structural measures and the method of sorting: In: Shepard, R. N., Romney, A. K. & Nerlove, S. B. (Eds). 225-249.
- Boorman, S. A. & Olivier, D. C. 1973. Metrics on spaces of finite trees. *Journal of Mathematical Psychology*, 10, 26-59.
- Boyce, D. E., Farhi, A. & Weischedel, R. 1974. Optimal subset selection. Multiple regression, interdependence and optimal network algorithms. Berlin: Springer.
- Boyd, J. P. & Wexler, K. N. 1973. Trees with structure. *Journal of Mathematical Psychology*, 10, 115-147.
- Bron, C. & Kerbasch, J. 1973. Algorithm 457. Finding all cliques of an undirected graph. *Communications of the ACM*, 16, 575-577.
- Brucker, P. 1978. On the complexity of clustering problems. In: Henn, R., Korte, B., & Oettli, W., 45-54.
- Büttner, J. 1975. Zur Clusteranalyse. *Biometrical Journal*, 17, 163-179.
- Bunemann, P. 1971. The recovery of trees from measures of dissimilarity. In: Hodson, F. R., Kendall, D. G. & Tautu, P. (Eds). 387-395.
- Carroll, J. D. 1976. Spatial, non-spatial and hybrid models for scaling. *Psychometrika*, 41, 439-463.
- Carroll, J. D. & Chang, J. J. 1973. A method for fitting a class of hierarchical tree structure models to dissimilarities and its application to some „Body Parts“ data of Miller's. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 8, 1097-1098.
- Garroll, J. D. & Pruzansky, S. 1980. Discrete and hybrid scaling models. In: Lantermann, E. D. & Feger, H. (Eds). 108-139.
- Christofides, N. 1975. Graph theory. An algorithmic approach. London: Academic Press.

- Christofides, N. 1975. Hamiltonian circuits and the travelling salesman problem: A survey. In: Roy, B. (Ed.). 149-171.
- Cohen, J. 1978. Multiple regression as a general data analytic system. *Psychological Bulletin*, 70, 426-433.
- Cole, A. J. (Ed.). 1969. *Numerical taxonomy*. London: Academic Press.
- Cole, A. J. & Wishart, D. 1970. An improved algorithm for the Jardine-Sibson method of generating overlapping clusters. *Computer Journal*, 13, 156-163.
- Colonus, H. & Schulze, H.-H. 1977. Trees constructed from empirical relations. *Braunschweiger Berichte aus dem Institut für Psychologie* 1.
- Colonus, H. & Schulze, H.-H. 1979. Repräsentation nichtnumerischer Ähnlichkeitsdaten durch Baumstrukturen. *Psychologische Beiträge*, 21, 98-111.
- Cormack, R. M. 1971. A review of classification. *Journal of the Royal Statistical Society (Series A)* 134, 321-367.
- Corsten, L. C. A. & Hermans, J. (Eds). 1978. *Compstat 1978. Proceedings in computational statistics*. Wien: Physica.
- Cunningham, J. P. 1974. Finding the optimal tree-realization of a proximity matrix. Paper presented of the Mathematical Psychology Meetings, Ann Arbor, August 1974.
- Cunningham, J. P. 1978. Free trees and bidirectional trees as representations of psychological distances. *Journal of Mathematical Psychology*, 17, 165-188.
- Cunningham, K. M. & Ogilvie, J. C. 1972. Evaluation of hierarchical grouping techniques: A preliminary study. *Computer Journal*, 15, 209-213.
- D'Andrade, R. 1978. U-statistic hierarchical clustering. *Psychometrika*, 43, 59-67.
- Day, W. H. E. 1977. Validity of clusters formed by graph-theoretic cluster methods. *Mathematical Biosciences*, 36, 299-317.
- Degerman, R. 1970. Multidimensional analysis of complex structure: Mixtures of class and quantitative Variation. *Psychometrika*, 35, 475-491.
- Degerman, R. 1972. The geometric representation of some simple structures. In: Shepard, R. N., Romney, A. K. & Nerlove, S. B. (Eds). 194-211.
- Dixon, L. C. W., Spedicato, E. & Szegő, G. P. (Eds). 1980. *Nonlinear optimization. Theory and algorithms*. Boston: Birkhäuser.
- Dixon, W. J. (Ed.). 1975. *BMDP. Biomedical computer programs*. Berkeley: University of California Press.
- Dixon, W. J. & Brown, M. B. 1977. *BMDP-77. Biomedical Computer programs. P-Series*. Berkeley: University of California Press, 418ff.
- Duran, B. S. & Odell, P. L. 1974. *Cluster analysis*. Berlin: Springer.
- Eckes, Th. & Roßbach, H. 1980. *Clusteranalysen*. Stuttgart: Kohlhammer.
- Edelbrock, C. 1979. Mixture model tests of hierarchical clustering algorithms: The problem of classifying everybody. *Multivariate Behavioral Research*, 14, 367-384.

- Edelbrock, C. & McLaughlin, B. 1980. Hierarchical cluster analysis using intraclass correlations: A mixture model study. *Multivariate Behavioral Research*, 15, 299-384.
- Edgington, E. S. 1969. *Statistical inference: The distribution-free approach*. New York: Mc Graw-Hill.
- Edgington, E. S. 1980. *Randomization tests*. New York: Marcel Dekker.
- Edwards, A. W. F. & Cavalli-Sforza, L. L. 1965. A method for cluster analysis. *Biometrics*, 21, 362-375.
- Eisler, H. 1973. The algebraic and statistical tractability of the city block metric. *British Journal of Mathematical and Statistical Psychology*, 26, 212-218.
- Enslein, K., Raiston, A. & Wilf, H. S. (Eds). 1977. *Statistical methods for digital Computers* (Vol. 3). New York: Wiley.
- Estabrook, G. F. (Ed.). 1975. *Proceedings of the Eighth International Conference on Numerical Taxonomy*. S. F.: Freeman.
- Eye, A. von 1977a. über die Verwendung von Quadriken zur einbeschreibenden Klassifikation. *Biometrical Journal*, 19, 283-290.
- Eye, A. von 1977b. Zum Vergleich zwischen der hierarchischen Clusteranalyse nach Ward und MACS, einer mehrdimensionalen automatischen Clustersuchstrategie. *Psychologische Beiträge*, 19, 201-217.
- Eye, A. von & Wirsing, M. 1978. An attempt for a mathematical foundation and evaluation of Macs, a method for multidimensional automatic cluster detection. *Biometrical Journal*, 20, 655-666.
- Farris, J. S. 1969. On the cophenetic correlation coefficient. *Systematic Zoology* 18, 279-285. In: Atchley, W. R. & Bryant, E. H. (Eds). 1975, 398-404.
- Fisher, W. D. 1958. On grouping for maximum homogeneity. *Journal of the American Statistical Association*, 53, 789-798.
- Flachsmeyer, J. 1977<sup>3</sup>. *Kombinatorik*. Berlin (Ost): VEB Deutscher Verlag der Wissenschaften.
- Frank, O. 1978. Inferences concerning cluster structure. In: Corsten, L. C. A. & Hermans, J. (Eds). 259-265.
- Friendly, M. 1977. In search of the M-Gram: The structure of organization in free recall. *Cognitive Psychology*, 9, 188-249.
- Furnival, G. M. & Wilson, R. W. 1974. Regression by leaps and bounds. *Technometrics*, 16, 499-511.
- Gower, J. C. 1975. Goodness-of-fit criteria for classification models and other patterned structures. In: Estabrook, G. F. (Ed.). 38-62.
- Gower, J. C. & Ross, G. J. S. 1969. Minimum spanning trees and single linkage cluster analysis. *Applied Statistics*, 18, 54-64.
- Harary, F. 1974. *Graphentheorie*. München: Oldenbourg.
- Hartigan, J. A. 1967. Representation of similarity matrices by trees. *Journal of the American Statistical Association*, 62, 1140-1158.

- Hartigan, J. A. 1972. Direct clustering of a data matrix. *Journal of the American Statistical Association*, 67, 123-129.
- Hartigan, J. A. 1975: Clustering algorithmus. New York: Wiley.
- Hartigan, J. A. 1977. Distribution problems in clustering. In: van Ryzin, J. (Ed.). 45-71.
- Hausmann, D. 1978. Integer programming and related areas. A classified bibliography, 1976-1978. Berlin: Springer.
- Henn, R., Korte, B. & Oettli, W. 1978. Optimization and Operations Research. Proceedings, Bonn 1977. Berlin: Springer.
- Herden, W. & Steinhausen, D. 1979. Ein Verfahren zur Berechnung eines globalen Minimums beim Varianzkriterium in der Cluster-Analyse. In: Späth, H. (Ed.). 115-142.
- Hodson, F. R., Kendall, D. G. & Tautu, P. 1971. Mathematics in the archaeological and historical sciences. Edinburgh University Press.
- Horst, R. 1979. Nichtlineare Optimierung. München: Hanser.
- Hubert, L. J. 1972. Some extensions of Johnson's hierarchical clustering algorithms. *Psychometrika*, 37, 261-274.
- Hubert, L. J. 1973a. Monotone invariant clustering procedures. *Psychometrika*, 38, 47-62.
- Hubert, L. J. 1973b. Min and max hierarchical clustering using asymmetric similarity measures. *Psychometrika*, 38, 63-72.
- Hubert, L. J. 1974a. Approximative evaluation techniques for the single-link and complete-link hierarchical clustering procedures. *Journal of the American Statistical Association*, 69, 698-704.
- Hubert, L. J. 1974b. Some applications of graph theory to clustering. *Psychometrika*, 39, 283-309.
- Hubert, L. J. 1974c. Spanning trees and aspects of clustering. *British Journal of Mathematical and Statistical Psychology*, 27, 19-88.
- Hubert, L. J. 1976. Seriation using asymmetric proximity measures. *British Journal of Mathematical and Statistical Psychology*, 29, 32-52.
- Hubert, L. J. 1977a. A set-theoretical approach to the problem of hierarchical clustering. *Journal of Mathematical Psychology*, 15, 70-88.
- Hubert, L.J. 1977b. Data analysis implications of some concepts related to the cuts of a graph. *Journal of Mathematical Psychology*, 15, 199-208.
- Hubert, L. J. 1978. Generalized proximity function comparison. *British Journal of Mathematical and Statistical Psychology*, 31, 179-192.
- Hubert, L. J. 1979a. Comparison of sequences. *Psychological Bulletin*, 86, 1098-1106.
- Hubert, L. J. 1979b. Matching models in the analysis of cross-classifications. *Psychometrika*, 44, 21-41.

- Hubert, L. J. 1979c. Generalized concordance. *Psychometrika*, 44, 135-142.
- Hubert, L. J. 1980. Analyzing proximity matrices: The assessment of internal variation in combinatorial structure. *Journal of Mathematical Psychology*, 21, 247-264.
- Hubert, L. J. & Baker, F. B. 1977a. An empirical comparison of baseline models for goodness-of-fit in r-diameter hierarchical clustering. In: Ryzin, J. van (Ed.). 131-153.
- Hubert, L. J. & Baker, F. B. 1977b. The comparison and fitting of given classification schemes. *Journal of Mathematical Psychology*, 16, 233-253.
- Hubert, L. J. & Baker, F. B. 1978a. Analyzing the multitrait-multimethod matrix. *Multivariate Behavioral Research*, 13, 163-179.
- Hubert, L. J. & Baker, F. B. 1978b. Applications of combinatorial programming to data analysis: The traveling salesman and related problems. *Psychometrika*, 43, 81-91.
- Hubert, L. J. & Baker, F. B. 1979. A note on analyzing the multitrait-multimethod matrix: An application of a generalized proximity function comparison. *British Journal of Mathematical and Statistical Psychology*, 32, 179-184.
- Hubert, L. J. & Schultz, J. V. 1975. Hierarchical clustering and the concept of space distortion. *British Journal of Mathematical and Statistical Psychology*, 28, 121-133.
- Hubert, L. J. & Schultz, J. 1976. Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 29, 190-241.
- Hull, C. H. & Nie, N. N. 1979. *SPSS-Update: New procedures and facilities for releases 7 and 8*. New York: Mc Graw-Hill.
- Jardine, N. & Sibson, R. 1968. The construction of hierarchic and non-hierarchic classifications. *Computer Journal*, 11, 177-184.
- Jardine, N. & Sibson, R. 1971. *Mathematical taxonomy*. New York: Wiley.
- Jöreskog, K. G. 1978. Structural analysis of covariance and correlation matrices. *Psychometrika*, 43, 443-477.
- Johnson, S. C. 1967. Hierarchical clustering schemes. *Psychometrika*, 32, 241-254.
- Kallina, H. 1967. Das Unbehagen in der Faktorenanalyse. *Psychologische Beiträge*, 10, 81-86.
- Kasting, C. 1976. *Integer programming and related areas. A classified bibliography*. Berlin: Springer.
- Kaufmann, A. 1975. *Introduction to the Theory of fuzzy subsets*. New York: Academic Press.
- Kennedy, W. J. & Gentle, J. E. 1980. *Statistical computing*. New York: Marcel Dekker Inc.
- Kirn, Ki Hang & Roush, F. W. 1978. Ultrametrics and matrix theory. *Journal of Mathematical Psychology*, 18, 195-203.

- Kleiter, E. F. 1980. Voraussetzungs-Koeffizient und Voraussetzungs-Clusteranalyse bei numerischen Daten. Teil I: Doppelte Abbildung und Voraussetzungs-Clusteranalyse. Teil II: Koeffizienten und Anwendungsbeispiele. Zeitschrift für erziehungswissenschaftliche Forschung, 14, 65-89 und 127-151.
- Kleiter, E. F. & Fillbrandt, H. 1973. Hypothesenorientierte Kategorienbildung und hierarchische Cluster-Analyse. Psychologische Beiträge, 15, 603-633.
- Kleiter, E. F. & Petermann, F. 1977. Abbildung von Lernwegen. München: Oldenbourg.
- Kleiter, E. F. & Timmermann, G. 1977. über einen Algorithmus zur hierarchischen Voraussetzungs-Struktur-Analyse. Psychologische Beiträge, 19, 355-390.
- Klix, F., Sydow, H. & Wysotzki, F. (Eds). 1974. Erkennungs- und Klassifizierungsprozesse. Berlin (Ost): VEB Verlag der Wissenschaft.
- Kopp, B. 1978a. Hierarchical classification I: Single-linkage method. Biometrical Journal, 20, 495-501.
- Kopp, B. 1978b. Hierarchical classification II: Complete-linkage method. Biometrical Journal, 20, 597-602.
- Kopp, B. 1978c. Hierarchical classification III: Average-linkage, median, centroid, WARD, flexible Strategie. Biometrical Journal, 20, 703-711.
- Krishnaiah, P. R. (Ed.). 1969. Multivariate analysis Vol. II. New York: Academic Press.
- Kruskal, J. 1977. The relationship between multidimensional scaling and clustering. In: Ryzin, J. van (Ed.). 17-44.
- Kruskal, J. & Carroll, J. B. 1969. Geometrical models and badness-of-fit functions. In: Krishnaiah, P. R. (Ed.). 639-671.
- Kubicki, St., Herrmann, W. M. & Laudahn, G. (Eds). 1980. Faktorenanalyse und Variablenbildung aus dem Elektroenzephalogramm. Stuttgart: Fischer.
- Kuiper, F. K. & Fisher, L. 1975. A Monte Carlo comparison of six clustering procedures. Biometrics, 31, 777-783.
- Lance, G. N. & Williams, W. T. 1967. A general theory of classificatory sorting strategies: I. Hierarchical Systems. Computer Journal, 9, 373-380.
- Lantermann, E. D. & Feger, H. (Eds). 1980. Similarity and choice. Papers in honour of Clyde Coombs. Bern: Huber.
- Lawler, E. L. 1975. The quadratic assignment problem: A brief review. In: Roy, B. (Ed.). 351-360.
- Lazarsfeld, P. F. & Henry, N. W. 1968. Latent structure analysis. Boston: Houghton, Mifflin.
- Lee, H. B. & Mac Queen, J. B. 1980. A K-means cluster analysis Computer programm with cross-tabulations and next-nearest-neighbor analysis. Educational and Psychological Measurement, 40, 130-138.
- Leuschner, D. 1973. Numerische Taxonomie und algebraische Struktur. Biometrical Journal, 15, 271-275.

- Lienert, G. A. 1978. Verteilungsfreie Methoden in der Biostatistik, Band II. Meisenheim: Hain.
- Ling, R. F. 1971. Cluster analysis. New Haven, Conn.: Dissertation, Yale University.
- Ling, R. F. 1972. On the theory and construction of K-clusters. *Computer Journal*, 15, 326-332.
- Ling, R. F. 1973. A probability theory of cluster analysis. *Journal of the American Statistical Association*, 68, 159-164.
- Ling, R. F. 1975. An exact probability distribution on the connectivity of random graphs. *Journal of Mathematical Psychology*, 12, 90-98.
- Ling, R. F. & Killough, G. G. 1976. Probability tables for cluster analysis based on a theory of random graphs. *Journal of the American Statistical Association*, 71, 293-300.
- Lingoes, J. C. 1973. The Guttman-Lingoes nonmetric program series. Ann Arbor: Mathesis Press.
- Lingoes, J. C. & Cooper, T. 1971. PEP-I: A Fortran IV (G) program for Guttman-Lingoes nonmetric probability clustering. *Behavioral Science*, 16, 259-261.
- Lingoes, J. C., Roskam, E. E. & Borg, J. 1976. Geometric representations of relational data structures. Ann Arbor: Mathesis Press.
- Marsten, R. E. 1975. An algorithm for large set partitioning problems. In: Roy, B. (Ed.). 259-267.
- Matejcek, M. & Schenk, G. K. 1976. Quantitative analysis of the EEG. Proceedings of 2nd Symposium of the Study Group for EEG Methodology, Jongny sur Vevey, Mai 1975; Konstanz: AEG-Telefunken.
- Matula, D. W. 1977. Graph theory techniques for cluster analysis algorithms. In: Ryzin, J. van (Ed.). 95-129.
- Mc Henry, C. E. 1978. Computation of a best subset in multivariate analysis. *Applied Statistics*, 27, 291-296.
- Mc Intyre, R. M & Blashfield, R. K. 1980. A nearest-centroid technique for evaluating the minimum-variance clustering procedure. *Multivariate Behavioral Research*, 15, 225-238.
- Mc Quitty, L. L. 1968. Multiple clusters, types, and dimensions from iterative intercolumnar correlational analysis. *Multivariate Behavioral Research*, 3, 465-477.
- Mc Quitty, L. L. & Clark, J. A. 1968. Clusters from iterative, intercolumnar correlational analysis. *Educational and Psychological Measurement*, 28.
- Milligan, G. W. 1979. Ultrametric hierarchical clustering algorithms. *Psychometrika*, 44, 343-346.
- Milligan, G. W. 1980. An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45, 325-342.
- Milligan, G. W. & Sokol, L. M. 1980. A two-stage clustering algorithm with robust recovery characteristics. *Educational and Psychological Measurement*, 40, 755-759.

- Mustonen, S. 1978. Digression analysis: Fitting alternative regression models to heterogenous data. In: Corsten, L. C. A. & Hermans, J. (Eds). 95-101.
- Müller-Merbach, H. 1970. Optimale Reihenfolgen. Berlin: Springer.
- Negoita, C. V. & Ralescu, D. A. 1975. Applications of fuzzy sets to systems analysis. Basel: Birkhäuser.
- Nenninger, P. 1980. Anwendungsmöglichkeiten der Graphentheorie in der Erziehungswissenschaft. Zeitschrift für Empirische Pädagogik, 4, 85-106.
- Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. & Bent, D. H. 1975<sup>2</sup>. SPSS - Statistical package for the social sciences. New York: Mc Graw-Hill.
- Oldenburger, H.-A. & Becker, D. 1976. Are there clusters of frequencies in power-spectra of EEG? How to find and prove them statistically. In: Matejcek, M. & Schenk, G. K. (Eds). 601-611.
- Oldenburger, H.-A. & Schwibbe, M. 1980. Konstruktive Kritik des Einsatzes dimensionaler Dekompositionsverfahren für EEG-Frequenzkomponenten. In: Kubicki, St., Herrmann, W. M. & Laudahn, G. (Eds). 47-60.
- Opitz, O. 1980. Numerische Taxonomie. Stuttgart: Fischer, UTB 918.
- Patrinos, A. N. & Hakimi, S. L. 1972. The distance matrix of a graph and its tree realization. Quarterly of Applied Mathematics, 30, 255-269.
- Peay, E. R. 1975. Nonmetric grouping: Clusters and cliques. Psychometrika, 40, 297-313.
- Rand, W. M. 1971. Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association, 66, 846-850.
- Rao, M. R. 1971. Cluster analysis and mathematical programming. Journal of the American Statistical Association, 66, 622-626.
- Revelle, W. 1978. INCLUST: A cluster analytic approach to exploratory and confirmatory scale construction. Behavioral Research Methods and Instrumentation, 10, 739-742.
- Revelle, W. 1979. Hierarchical cluster analysis and the internal structure of tests. Multivariate Behavioral Research, 14, 57-74.
- Rohlf, F. G. 1974. Graphs implied by the Jardine-Sibson overlapping clustering methods,  $B_k$ . Journal of the American Statistical Association, 69, 705-710.
- Rohlf, F. J. 1975. A new approach to the computation of the Jardine-Sibson  $B_k$  clusters. Computer Journal, 18, 164-168.
- Rasch, E. 1975. Cognitive reference points. Cognitive Psychology, 7, 532-547.
- Rasch, E. 1978. Principles of categorization. In: Rasch, E. & Lloyd, B. B. (Eds). 27-48.
- Rasch, E. & Lloyd, B. B. (Eds). 1978. Cognition and categorization. Hillsdale, N. J.: Erlbaum.
- Roskam, E. E. 1975. Nonmetric data analysis. Department of Psychology, University of Nijmegen, Holland, Report 75-MA-13.

- Roy, B. (Ed.). 1975. Combinatorial programming: Methods and applications. Proceedings of the NATO Advanced Study. Institut Dordrecht (Holland); D. Reidel Publ. Comp., 1975.
- Ryzin, J. van (Ed.). 1977. Classification and clustering. New York: Academic Press.
- Sattath, S. & Tversky, A. 1977. Additive similarity trees. *Psychometrika*, 42, 319-345.
- Schultz, J. V. & Hubert, L. 1973. Data analysis and the connectivity of random graphs. *Journal of Mathematical Psychology*, 10, 421-428.
- Shepard, R. N. & Arabie, Ph. 1979. Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86, 87-123.
- Shepard, R. N., Romney, A. K. & Nerlove, S. B. (Eds). 1972. Multidimensional scaling. Theory and application in the behavioral sciences. Vol. 1. New York: Seminar Press.
- Silverstein, A. B. 1980. Cluster analysis of the Wechsler intelligence scale for children-revised. *Educational and Psychological Measurement*, 40, 51-54.
- Sneath, P. H. A. 1969. Evaluation of clustering methods. In: Cole, A. J. (Ed.). 257-271.
- Sneath, P. H. A. & Sokal, R. R. 1973. Numerical taxonomy. S. F.: Freeman.
- Sodeur, W. 1976. Buchbesprechung zu Vogel 1975. *Allgemeines Statistisches Archiv*, 415-416.
- Sokal, R. R. & Rohlf, F. J. 1962. The comparison of dendrograms by objective methods. *Taxon*, 11, 33-39.
- Sokal, R. R. & Sneath, P. H. A. 1963. Principles of numerical taxonomy. S. F.: Freeman.
- Späth, H. 1975. Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion. München: Oldenbourg.
- Späth, H. 1976.  $L_1$  Cluster Analyse. *Computing*, 16, 379-387.
- Späth, H. 1977. Fallstudien Cluster-Analyse. München: Oldenbourg.
- Späth, H. 1977a. Partitionierende Cluster-Analyse für große Objektmengen mit binären Merkmalen am Beispiel von Firmen und deren Berufsgruppenbedarf. In: Späth, H., 1977, 63-80.
- Späth, H. 1977b. Computational experiences with the exchange method applied to four commonly used partitioning cluster analysis criteria. *European Journal of Operation Research*, 1, 23-31.
- Späth, H. (Ed.). 1979. Ausgewählte Operations Research Software in FORTRAN. München: Oldenbourg.
- Späth, H. & Müller, R. 1979. Das Austauschverfahren für die skalierungsvariante Methode der adaptiven Distanzen in der Cluster-Analyse. In: Späth, H. (Ed.). 143-163.

- Steiger, J. H. 1980: Testing pattern hypotheses on correlation matrices: Alternative statistics and some empirical results. *Multivariate Behavioral Research*, 15, 335-352.
- Steinhausen, D. & Langer, K. 1977. *Clusteranalyse. Einführung in Methoden und Verfahren der automatischen Klassifikation*. Berlin: de Gruyter.
- Tinhofer, G. 1980. *Zufallsgraphen*. München: Hanser.
- Tyron, R. C. 1939. *Cluster analysis*. Ann Arbor, Mich.: Edwards Brothers.
- Tukey, J. W. 1977. *Exploratory data analysis*. Reading, Mass.: Addison-Wesley.
- Tversky, A. 1977. Features of similarity. *Psychological Review*, 84, 327-352; auch in: Lantermann, E. D. & Feger, H. (Eds). 1980, 12-53.
- Tversky, A. & Sattath, S. 1979. Preference trees. *Psychological Review*, 86, 542-573.
- Tzeng, O. C. S. & May, W. H. 1979. On rotation of a Johnson hierarchical tree structure. *Educational and Psychological Measurement*, 39, 733-741.
- Vogel, F. 1975. *Probleme und Verfahren der numerischen Klassifikation*. Göttingen: Vandenhoeck & Ruprecht.
- Ward, J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236-244.
- White, H. C. 1977. Probabilities of homomorphic mappings from multiple graphs. *Journal of Mathematical Psychology*, 16, 121-134.
- Williams, W. T. & Lance, G. N. 1977. Hierarchical classificatory methods. In: Enslein, K., Ralston, A. & Wilf, H. S. (Eds). 269-295.
- Wishart, D. 1978. *CLUSTAN*. User manual, Program library unit, Edinburgh University. (Program Library Unit, Edinburgh University 18 Buccleuch Place, Edinburgh EH8 9LN Scotland).
- Welfe, J. H. 1970. Pattern clustering by multivariate mixture analysis. *Multivariate Behavioral Research*, 5, 329-350.
- Wolfe, J. H. 1978. Comparative cluster analysis of patterns of vocational interest. *Multivariate Behavioral Research*, 13, 33-44.
- Woodward, J. A. & Bentler, P. M. 1979. Application of optimal sign-vectors to reliability and cluster analysis. *Psychometrika*, 44, 337-342.
- Wright, W. E. 1974. An axiomatic specification of euclidean analysis. *Computer Journal*, 17, 355-364.
- Yahil, A. & Brown, M. B. 1976. On separating clusters from background. *Technometrics*, 18, 55-58.
- Yeh, R. T. & Bang, S. Y. 1975. Fuzzy relations, fuzzy graphs, and their applications to clustering analysis. In: Zadeh, L. A., Fu, King-Sun, Tanaka, K. & Shimura, M. (Eds). 125-149.
- Young, F. W., de Leeuw, J. & Takane, Y. 1980. Quantifying qualitative data. In: Lantermann, E. D. & Feger, H. (Eds). 150-179.

- Zadeh, L. A. 1977. Fuzzy sets and their application to pattern classification and clustering analysis. In: van Ryzin, J. (Ed.). 251-299.
- Zadeh, L. A., Fu, King-Sun, Tanaka, K. & Shimura, M. (Eds). 1975. Fuzzy sets and their applications to cognitive and decision Prozesses. New York: Academic Press.
- Lawrence Hubert danke ich für die Durchsicht der Druckvorlage dieses Artikels und seinen ergänzenden Hinweis auf das folgende wichtige Review, das leider nicht mehr eingearbeitet werden konnte:
- Dubes, R. & Jain, A. K. 1980. Clustering Methodologies in Exploratory Data Analysis. In: Yovits, M. C. Advances in Computers, Vol. 19, New York: Academic Press, 113-228.