# ALGORITMOS PARA DETERMINAR CONGLOMERADOS DE NIVEL Y CONGLOMERADOS ESTRATIFICADOS.

Por Miguel Sánchez García. Departamento de Estadística e I.O. Facultad de Matemáticas. Madrid.

## 1. Introducción.

El gran impulso que está recibiendo la investigación en campos tales como la sociología, psicología, medicina, biología, agricultura, etc.; donde las variables en estudio son en numerosas ocasiones cualitativas, e incluso cuando son cuantitativas no se ajustan con fiabilidad, en un gran número de casos, a las hipótesis de los modelos estadísticos lineales, tradicionalmente empleados en su tratamiento, ha motivado la búsqueda de otros métodos para tratar esta información, siendo uno de estos el análisis de conglomerados ó clustering, herramienta básica para las técnicas de ordenación y clasificación.

El objeto fundamental de estas técnicas consiste en determinar estructuras básicas de los sistemas multivariables. El estudio comienza con un universo simple compuesto de individuos. Estos individuos son caracterizados de acuerdo con un conjunto de parámetros (atributos, propiedades, caracteres, etc.) y son representados en un espacio multidimensional cuyos ejes coordenados son los parámetros.

Las técnicas de conglomerados tratan de agrupar los individuos de acuerdo con los valores de sus parámetros. Las agrupaciones mas comunmente utilizadas hasta el momento, han sido las jerarquicas, creemos que por su facilidad de computación; pero estimamos que darán mucha más luz, para la comprensión de determinados fenómenos, las no jerarquicas.

En el presente artículo, se dan dos algoritmos para computar esta 2ª clase de conglomerados, el primero de ellos calcula conglomerados de nivel; y el 2º calcula los conglomerados de un nivel superior, teniendo como base los conglomerados de un nivel dado.

## 2. Grafos y Conglomerados. Sus relaciones.

En este epígrafe demostraremos que para la obtención de conglomerados de nivel, son equivalentes los conceptos de grafo de proximidad completo y de clase de coeficientes de desemejanza.

Un grafo $G = (V,E)$ está compuesto por un conjunto $V$ finito y no vacío de vértices $\{0_1, 0_2, \ldots, 0_n\}$; y un conjunto $E = \{l_1, l_2, \ldots, l_m\}$ de aristas, con $m \leq n(n+1)/2$, donde cada arista $l_k$ está formada por un par de vértices $0_i, 0_j \in V$.

Si $m = n(n+1)/2$ el grafo se llama completo, y si los elementos de $E$ están ordenados, es decir $l_1 \leq l_2 \leq \ldots \leq l_m$, entonces el grafo se llama ordenado.

Un grafo de proximidad $G = (V,E)$ es un grafo ordenado, para el cual $V = \{0_1, 0_2, \ldots, 0_n\}$ es el conjunto de objetos que se han de agrupar, y $E = \{l_1, l_2, \ldots, l_m\}$ es el conjunto de aristas del grafo, de tal forma que la relación de orden entre las aristas, corresponda a una función de proximidad entre los objetos.

Un grafo de proximidad completo, es un grafo completo para el cual $(0_i, 0_i) = (0_j, 0_j) \leq (0_i, 0_j)$ para todo $0_i, 0_j \in V$.

En otro orden de ideas, sea $V$ un conjunto de $n$ objetos. Un coeficiente de desemejanza sobre $V$, es una aplicación

$$d : V \times V \longrightarrow R^+$$

que verifica las propiedades siguientes:

1ª. Para todo $0_i \in V$ $\quad d(0_i, 0_i) =$ cero

2ª. Para todo $0_i, 0_j \in V$, $\quad d(0_i, 0_j) = d(0_j, 0_i)$

De la definición se deduce que un coeficiente de desemejanza quede determinado cuando se conocen el conjunto de $n(n-1)/2$ números reales $\{d(0_i, 0_j)\}$ $\quad i = 1,2,\ldots,n-1, \quad j = i+1, i+2, \ldots, n.$

A continuación daremos el concepto de clase de coeficientes de desemejanza.

Diremos que dos coeficientes de desemejanza, $d_1$ y $d_2$, están relacionados, y lo denotaremos por $d_1 R d_2$, si existe una función

$$f ; R^+ \longrightarrow R^+$$

contínua y monótona creciente estrictamente, tal que para todo $O_i, O_j \in V$ se cumpla:

$$d_1(O_i,O_j) = f\, d_2(O_i,O_j)$$

Evidentemente la relación R es de equivalencia, y a cada clase de equivalencia se la denomina clase de coeficientes de desemejanza.

Teorema. $d_1 \ R \ d_2 \iff (d_1(O_i,O_j) \leq d_1(O_e,O_k) \iff d_2(O_i,O_j) \leq$ $\leq d_2(O_e,O_k))$.

Demostración.

$\Longrightarrow$ Si $d_1 R d_2$, existe $f : R^+ \longrightarrow R^+$ monótona creciente estrictamente, tal que $d_1(O_i,O_j) = f\, d_2(O_i,O_j)$; entonces

$$d_1(O_i,O_j) \leq d_1(O_e,O_k) \iff f(d_2(O_i,O_j)) \leq f(d_2(O_e,O_k)) \iff$$

$$\iff d_2(O_i,O_j) \leq d_2(O_e,O_k)$$

$\Longleftarrow$ De $d_1(O_i,O_j) \leq d_1(O_e,O_k) \iff d_2(O_i,O_j) \leq d_2(O_e,O_k)$ se sigue que $d_1(O_i,O_j) = d_1(O_e,O_k) \iff d_2(O_i,O_j) = d_2(O_e,O_k)$, como consecuencia sean:

$$0 < l_1^1 < l_1^2 < \ldots\ldots < l_1^k \qquad 0 < l_2^2 < l_2^2 < \ldots. < l_2^k$$

los valores de desemejanza distintos de $d_1$ y $d_2$ y sea f definida por:

$$f(l) = l . l_1^1 | l_2^1 \qquad \text{para} \qquad 0 \leq l \leq l_2^1$$

$$f(l) = l_2^j \frac{l - l_2^j}{l_2^{j+1} - l_1^j} (l_2^{j+1} - l_2^j) \qquad \begin{array}{l} j = 1,2,\ldots.,k-1 \\ l_2^j \leq l \leq l_2^{j+1} \end{array}$$

$$f(l) = l_1^k + (l - l_2^k) \qquad \text{para} \qquad l_2^k \leq l < \infty$$

Evidentemente f es una función continua y monótona creciente estrictamente tal que $d_2(0_i,0_j) = f_2(d_2(0_i,0_j))$ para todo $(0_i,0_j) \in V$, por tanto $d_1 \; R \; d_2$. c.q.d.

Dado un conjunto finito V y un coeficiente de desemejanza d, se puede construir un grafo de proximidad completo, colocando como longitud de la arista $(0_i,0_j)$ el valor $d(0_i,0_j)$, y ordenando las aristas por sus longitudes se obtiene el grafo deseado. Evidentemente cualquier coeficiente de desemejanza $d_1$ equivalente a d, da lugar al mismo grafo de proximidad.

Reciprocamente, dado un grafo de proximidad, se puede construir una clase de coeficientes de desemejanza, para lo cual basta asignar a las aristas $(0_i,0_i)$ el valor cero y a las aristas $(0_i,0_j)$ cualquier número real que conserve el orden estricto dado por el grafo de proximidad.

Mediante el razonamiento anterior se ha comprobado que los conceptos de grafo de proximidad completo y clase de coeficientes de desemejanza son equivalentes, en el sentido de que dado uno de ellos se puede construir el otro, y debido a ello los conceptos dados a continuación están referidos a ambos, aunque se enunciarán para grafos de proximidad.

Def. Los niveles de salto de un grafo de proximidad $P = (V,E)$, son los niveles $S = 0$, $S = m = |E|$ y todos los $S$, $1 \leq S \leq m-1$ para los cuales $l_S < l_{S+1}$. Para cada nivel de salto, el grafo $P_S = (V,E_S)$ donde $E_S = (l_1,l_2,...,l_S)$, con la relación de orden de E restringida a $E_S$, se denomina subgrafo de proximidad de orden S-ésimo de P.

Se denomina subgrafo umbral S-ésimo de P, al subgrafo $T_S = (V, E_S)$ donde $E_S$ se supone que no está ordenado.

<u>Def</u>. Un <u>conglomerado de nivel</u> del conjunto de objetos V, $V = \{0_1, 0_2, \ldots, 0_n\}$, es una familia L de subconjuntos no anidados de V, llamados conglomerados, que cubren V, es decir:

$$C, C' \in L \quad y \quad C \subset C' \Longrightarrow C = C' \quad \text{y además} \quad \bigcup_{C \in L} C = V.$$

<u>Def</u>. Un ciclado de un grafo $P = (V,E)$, es un subconjunto $A \subset V$, tal que para todo $(0_i, 0_j) \in A$, $(0_i, 0_j) \in E$.

<u>Def</u>. Se llama <u>conglomerado de nivel $l_S$</u>, para el grafo de proximidad $P = (V,E)$, a la familia $L_S$ de ciclados maximales del subgrafo umbral S-simo de $P = (V,E)$.

<u>Def</u>. Un conglomerado estratificado del conjunto de objetos V, $V = \{0_1, 0_2, \ldots, 0_n\}$, es una sucesión S de conglomerados de nivel de V, $S = (L_o, L_1, L_2, \ldots, L_k)$ tal que i) $L_o = \{(0_1), (0_2), \ldots$ $\ldots, (0_n)\}$ y ii) $\forall\, C \in L_i$ existe $C' \in L_{i+1}$ con $C \subset C'$.

<u>Def</u>. El conglomerado estratificado $S = \{L_o, L_1, \ldots, L_k\}$ es jerarquico, si cada conglomerado de nivel constituye una partición de E.

<u>Def</u>. Un conglomerado estratificado del grafo de proximidad completo $P = (V,E)$, viene dada por la sucesión $S = \{L_o, L_1, L_2, \ldots, L_k\}$, donde $L_i$, $i=0,1,2,\ldots,k$, es el conglomerado de nivel del grafo $P = (V,E)$.

<u>Nota</u>. Hemos supuesto que en el grafo de proximidad P, se identifican dos objetos $0_i$ $0_j$ si $(0_i, 0_i) = (0_i, 0_j)$ respecto de la relación de orden.

## 6. Algoritmo para hallar el conglomerado de nivel $l_S$ de un grafo de proximidad.

Este algoritmo construye el conglomerado de nivel $l_S$ sobre el número de objetos de V.

Es evidente que si $V = \{0_1\}$ entonces $L_S^1 = \{0_1\}$.
Supongamos que tenemos construidos el conglomerado de nivel $l_s$ para $V_k = \{0_1, 0_2, \ldots, 0_k\}$ y le denotamos por $L_s^k$, y queremos determinar el de $V_{k+1} = \{0_1, 0_2, \ldots, 0_k, 0_{k+1}\}$.

Sean $A = \{0/0 \in V_k$ y $(0, 0_{k+1}) \leq l_s\}$, $L_s^k = \{A_1, A_2, \ldots, A_m\}$,

$F = \{A_1', A_2', \ldots, A_m'\}$ con $A_j' = (A_j \cap A) \cup \{0_{k+2}\}$ y

$G = \{A_1, A_2, \ldots, A_m, A_1', A_2', \ldots, A_m'\}$

<u>Teorema</u>. Sea H la familia de elementos maximales de G respecto de la relación de inclusión, H entonces coincide con el conglomerado de nivel $l_s$ del grafo de proximidad $P_{k+1}$. Al conglomerado de nivel $l_s$ de $P_{k+1}$ le denotaremos por $L_s^{k+1}$.

<u>Demostración</u>. a) Sea $B \in L_s^{k+1}$, pueden suceder dos casos:

i) $0_{k+1} \notin B$, entonces $B \in L_s^k$ y por tanto existe $A_j$, maximal en G tal que $B = A_j$, luego $B \in H$.

ii) $0_{k+1} \in B$, entonces $B' = B - \{0_{k+1}\} \subset A$, existiendo $A_j \in L_s^k$ con $B' \subset A_j$, como consecuencia $B' \subset A_j \cap A$, pero si el contenido fuera estricto, se verificaría que $B \subset A_j'$ estrictamente y por tanto B no sería maximal, luego $B = A_j'$, y por ser maximal se deduce que $B \in H$ y como consecuencia de i) y ii) $L_s^{k+1} \subset H$.

b) De ser todos los elementos de H ciclados maximales se sigue la igualdad entre H y $L_s^{k+1}$. c.q.d.

Algoritmo

Paso 1. Se calcula $A = \{0/0 \in V_k \ \text{ y } \ (0,0_{k+1}) \leq 1_s\}$ y se coloca $J = \emptyset$.

Paso 2. Para $i=1,2,\ldots,m$ se calculan $B'_i = A \cap A_i$.

Si $B'_i = A_i$ se sustituye $A_i$ por $A'_i = A_i \cup \{0_{k+1}\}$ y se coloca $J = J \cup \{i\}$. Se coloca $J'' = \{\emptyset\}$

Paso 3. Respecto de la relación de inclusión, se ordenan los elementos $B'_i$ de mayor a menor, y se calcula el conjunto $J'$ formado por los nuevos índices de los conjuntos $B'_j$ para $j \in J$, colocando además $H = \{B'_1\}$ donde $B'_1$ es el primer conjunto después de la ordenación y se coloca $I = 1$.

Nota. En el paso siguiente, el conjunto $B'_i$ es el que ocupa el lugar $i$ después de la ordenación.

Paso 4. Para $i=2,3,\ldots,m$, se comprueba si $B'_i \subset B'_j$ para algún $B'_j \in H$. Si la inclusión es cierta se elimina a $B'_i$, en caso contrario se coloca $H = H \cup \{B'_i\}$ y se coloca $I = I+1$.

Si $i \in J'$, se coloca $J'' = J'' \cup \{I\}$ en otro caso se coloca $J'' = J''$ y se pone $k = 0$.

Paso 5. Para $i=1,2,\ldots,I$, si $i \in J''$ se coloca $H = H - \{B'_i\}$, en caso contrario se coloca $H = H$ y $k = k+1$.

Paso 6. Para $i=1,2,\ldots,k$ se coloca $Am+i = B'_i \cup \{0_{k+1}\}$

Paso 7. Se coloca $L_s^{k+1} = \{A_1, A_2, \ldots, A_m, A_{m+1}, \ldots, A_{m+k}\}$

Ejemplo.

Sea la matriz de desemejanza siguiente:

| A | B | C | D | E | F | G | H | I | J | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 6 | 6'5 | 2'5 | 2 | 1'5 | 3'5 | 1 | 0'5 | 2 | A |
| | | 6 | 4 | 6 | 11 | 9'3 | 4 | 9 | 8 | B |
| | | | 3 | 12 | 3'7 | 9 | 3 | 3 | 7 | C |
| | | | | 7 | 4'2 | 6 | 5 | 7 | 4 | D |
| | | | | | 8'9 | 4 | 7 | 3 | 9 | E |
| | | | | | | 7 | 8 | 5 | 13 | F |
| | | | | | | | 6 | 4 | 12 | G |
| | | | | | | | | 3 | 8 | H |
| | | | | | | | | | 4'5 | I |
| | | | | | | | | | | J |

La matriz del grafo para un umbral 6 es,

| A | B | C | D | E | F | G | H | I | J | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | A |
| | | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | B |
| | | | 1 | 0 | 1 | 0 | 1 | 1 | 0 | C |
| | | | | 0 | 1 | 1 | 1 | 0 | 1 | D |
| | | | | | 0 | 1 | 0 | 1 | 0 | E |
| | | | | | | 0 | 0 | 1 | 0 | F |
| | | | | | | | 1 | 0 | 0 | G |
| | | | | | | | | 1 | 0 | H |
| | | | | | | | | | 1 | I |
| | | | | | | | | | | J |

<u>Ciclo 1 K = 1</u> $L_6^1 = \{A\}$

<u>Ciclo 2 K = 2</u> $a = \{A\}$ y $L_6^2 = \{AB\}$

<u>Ciclo 3 K = 3</u> $a = \{B\}$ y $L_6^3 = \{AB, BC\}$

Ciclo 4 K = 4 $a = \{ABC\}$ $A'_1 = \{AB\}$ $A'_2 = \{BC\}$

$L_6^4 = \{ABD,\ BCD\}$

Ciclo 5 K = 5 $a = \{AB\}$ $A'_1 = \{AB\}$ $A'_2 = \{B\}$

$L_6^5 = \{ABD,\ BCD,\ ABE\}$

Ciclo 6 K = 6 $a = \{ACD\}$ $A'_1 = \{AD\}$ $A'_2 = \{CD\}$ $A'_3 = \{A\}$

$L_6^6 = \{ABD,\ BCD,\ ABE,\ ADF,\ CDF\}$

Ciclo 7 K = 7 $a = \{ADE\}$ $A'_1 = \{AD\}$ $A'_2 = \{D\}$ $A'_3 = \{AE\}$

$A'_4 = \{A\}$ $A'_5 = \{D\}$

$L_6^7 = \{ABD,\ BCD,\ ABE,\ ADF,\ CDF,\ ADG,\ AEG\}$

Ciclo 8 K = 8 $a = \{ABCDG\}$ $A'_1 = \{ABD\}$ $A'_2 = \{BCD\}$

$A'_3 = \{AB\}$ $A'_4 = \{AD\}$ $A'_5 = \{CD\}$

$A'_6 = \{ADG\}$ $A'_7 = \{AG\}$

$L_6^8 = \{ABDH,\ BCDH,\ ABE,\ ADF,\ CDF,\ ADGH,\ AEG\}$

Ciclo 9 K = 9 $a = \{ACEFGH\}$ $A'_1 = \{AH\}$ $A'_2 = \{CH\}$

$A'_3 = \{AE\}$ $A'_4 = \{AF\}$ $A'_5 = \{CF\}$

$A'_6 = \{AGH\}$ $A'_7 = \{AEG\}$

$L_6^9 = \{ABDH,\ BCDH,\ ABE,\ ADF,\ CDF,\ ADGH,\ AEG,\ CHI,\ AFI,\ CFI,\ AGHI, AEGI\}$

Ciclo 10 K = 10 $a = \{ADI\}$ $A'_2 = \{AD\}$ $A'_2 = \{D\}$ $A'_3 = \{A\}$

$A'_u = \{AD\}$ $A'_5 = \{D\}$ $A'_6 = \{AD\}$

$A'_7 = \{AI\}$ $A'_8 = \{I\}$ $A'_9 = \{AI\}$

$A'_{10} = \{I\}$ $A'_{11} = \{AI\}$

$L_6^{10} = \{ABDH,BCDH,ABE,ADF,CDF,ADGH,AEG,CHI,AFI,CFI,AGHI,AEGI,ADJ,AIJ\}$

Comentario. Es conveniente para este algoritmo que los objetos estén en orden decreciente del número de aristas que en ellos coinciden.

4. Algoritmo para hallar $L_{k+1}$ a partir de $L_k$.

Sea $L_k = \{A_1, A_2, \ldots, A_m\}$ el conglomerado de nivel $l_k$, y supongamos que el nivel $l_{k+1}$ añade al grafo $(V, E_k)$ la arista $(a,b)$. No hay pérdida de generalidad al añadir una sola arista, pues si el nivel $l_{k+2}$ añadiera al grafo más de una arista, se podría añadir subniveles entre $l_k$ y $l_{k+1}$, de forma tal que para cada subnivel solo se añadiera al grafo una arista.

Sea $A = \{0/0 \in V - \{a,b\}$ y $(0,a) \leq l_{k+1}$ y $(0,b) \leq l_{k+1}\}$, sea

$F = \{A'_1, A'_2, \ldots, A'_m\}$, donde $A'_i = (A_i \cap A) \cup \{a,b\}$

$G = \{A_1, A_2, \ldots, A_m, A'_1, A'_2, \ldots, A'_m\}$

Teorema. Sea $H$ la familia de elementos maximales de $G$ respecto de la relación de inclusión, $H$ entonces coincide con el conglomerado de nivel $l_{k+1}$, del grafo de proximidad $P_{k+1} = (V, E_{k+1})$. Al conglomerado de nivel $l_{k+1}$ de $P_{k+1}$ le denotaremos por $L_{k+1}$.

Demostración. a) Sea $B \in L_{k+1}$, pueden entonces suceder dos casos:

i) $\{a,b\} \not\subset B$, entonces $B \in L_k$ y es elemento maximal de $G$, por tanto $B \in H$.

ii) $\{a,b\} \subset B$, sea $B' = B - \{a,b\}$ y evidentemente $B' \subset A$, y además existe $j$ tal que $B' \subset A_j$, luego $B' \subset A_j \cap A$, pero por ser $B$ maximal $B' = A_j \cap A$ y $B = A'_j$ maximal en $H$, por tanto $L_{k+1} \subset H$.

b) Pero como $L_{k+1}$ tiene todos los ciclados maximales, y H solo contiene ciclados maximales se tiene que $H = L_{k+1}$ c.q.d.

Algoritmo

Paso 1. Se calcula $A = \left\{ 0/0 \in V - \{a,b\},\ (0,a) \leq 1_{k+1}, (0,b) \leq \leq 1_{k+1} \right\}$, colocando $J = \emptyset$, $I = \emptyset$.

Paso 2. Para $i = 1,2,\ldots,m$ se calculan $B_i' = A \cap A_i$

Si $B_i' = A_i$ y $a \in A_i$ se coloca $I = I \cup \{i\}$

Si $B_i' = A_i$ y $b \in A_i$ se coloca $J = J \cup \{i\}$

Paso 3. Respecto de la relación de inclusión, se ordenan los elementos $B_i'$ de mayor a menor, y se calculan los conjuntos I' y J' formados por los nuevos índices de los conjuntos $B_i'$ para $i \in I$, $B_i'$ para $i \in J$, colocando $H = \{B_1'\}$, donde $B_1'$ es el primer conjunto después de la ordenación y se coloca $i_o = 1$.

Nota. En el paso siguiente, el conjunto $B_i'$ es el que ocupa el lugar i después de la ordenación.

Paso 4. Para $i=2,3,\ldots,m$; se comprueba si $B_i' \subset B_j'$ para algún $B_j' \in H$. Caso afirmativo se elimina $B_i'$. Si además $i \in I'$, y es $j \in I$ el índice que en el paso 3 se convierte en I, entonces también se elimina $A_j \in L_k$. Igual proceso se sigue si $i \in J'$. Si $B_i' \not\subset B_j'$ para ningún $B_j' \in H$, se coloca $H = H \cup \{B_j'\}$ e $i_o = i_o + 1$.

Paso 5. Para cada $B_i' \in H$ si $i \in I'$, se sustituye el $A_j \in L_k$ tal que $A_j = B_i'$ por $A_j = B_i' \cup \{a,b\}$ y lo mismo se hace para $i \in J'$, en ambos casos se retira $B_i'$ de H.

Paso 6. Para cada $B'_i \in H$ se hace $B_i = B'_i \cup \{a,b\}$ y se hace que $B \in S$.

Los elementos de $S$, más los que quedan en $L_k$ son los elementos de $L_{k+1}$.

Ejemplo. En el ejemplo anterior el umbral siguiente corresponde a un valor de 6'5, que hace que se una el grafo el par (AC).

Sea $A' = \{0/d(0,A) \leq 6'5\ d(0,C) \leq 6'5,\ 0 \neq A\ 0 \neq C\}$ entonces $A' = \{BDFHI\}$

El conjunto $L_k$ viene dado por

$L_6$ = {ABDH, BCDH, ABE, ADF, CDF, ADGH, AEG, CHI, AFI, CFI, AGHI, AEGI, ADJ, AIJ}

Las intersecciones de los elementos de $L_6$ con $A'$ son:

BDH, BDH, B, DF, DF, DH, HI, FI, FI, HI, I, D, I

De estas intersecciones se obtiene

$L_7$ = {ABCDH, ABE, ACDF, ADGH, ACDH, ACHI, ACFI, AGHI, AEGI, ADJ, AIJ}

Comentario. En el paso 4, se quitarían BDH, que a su vez eliminaria a BCDH; B; DF que a su vez eliminaría a CDF; DH; FI que a su vez eliminaría CFI; HI; I; D; I; con lo que no quedarían los siguientes elementos:

BDH, DF, HI, FI

BDH se uniría con AC y ABCDH sustituiría ABDH

ACDF sustituiría a ADF

ACFI sustituiría a AFI, y se añadiría ACHI

Bibliografía.

Jardine, N.- Sibson, R.: "Mathematical Taxonomy". (1971). John Wiley.

Sanchez, M.- "Conglomerados y sus aplicaciones". Trabajos de Estadística y de I.O.; vol. XXVII, 1976, pp. 159-173.

Vam Ryzin, J. "Classification and Clustering". (1977). Academic Press.