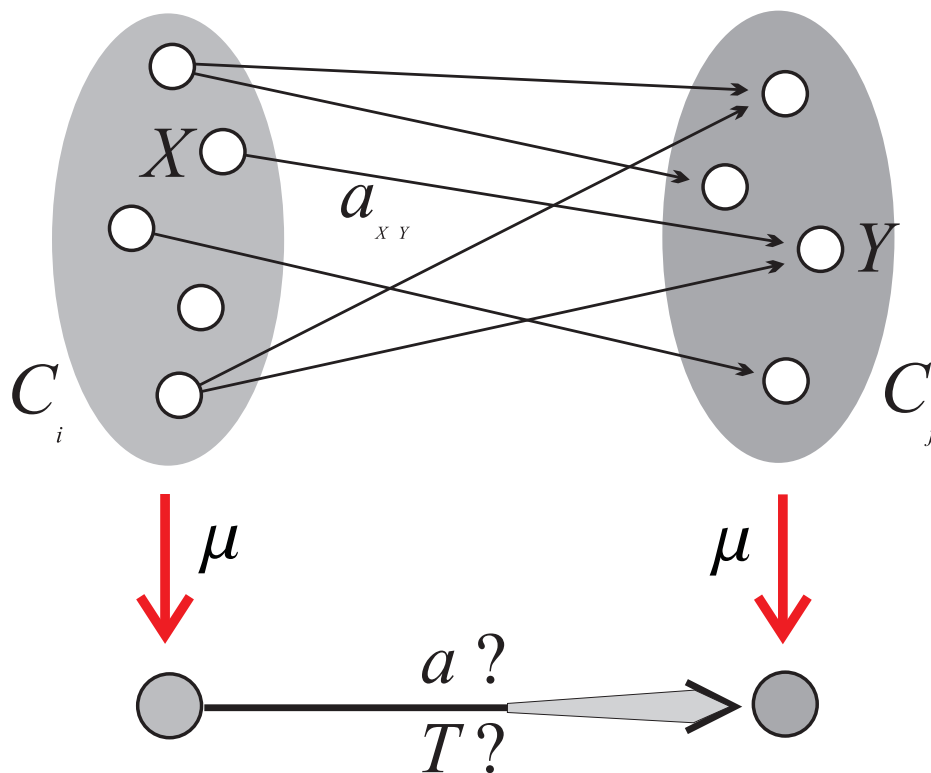


## Uvod v bločne modele

Cilj bločnega modeliranja je, da poskušamo večje, nepregledno omrežje skrčiti glede na predpostavljeno vrsto enakovrednosti na manjše omrežje, kjer so enote skupine enakovrednih enot. Tako dobljena struktura (relacija, matrika, graf) je preglednejša in ustrežnejša za interpretacije.



Slika 1: Bločno modeliranje.

Tabela 1: Sampsonovi podatki – naklonjenost

			1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8
John Bosco	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0
Gregory	2	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0
Basil	3	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1
Peter	4	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0
Bonaventur	5	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
Berthold	6	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Mark	7	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
Victor	8	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
Ambrose	9	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0
Romuald	10	0	0	0	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
Louis	11	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0
Winfriid	12	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Amand	13	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1
Hugh	14	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0
Boniface	15	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
Albert	16	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
Elias	17	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Simplicius	18	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0

V danem omrežju želimo določiti skupine enot, ki imajo enako (podobno) vlogo - imajo enak (podoben) vzorec povezanosti z drugimi enotami. Te skupine sestavljajo razvrstitev. Bločni model tedaj sestavljajo strukture (relacije, matrike, grafi), ki jih dobimo, ko vse enote iz iste skupine stisnemo v eno.

V primeru Sampsonovega omrežja menihov (naklonjenost v četrtem obdobju) so Breiger, Boorman in Arabie (1975) so z njihovim algoritmom CONCOR ugotovili, da omrežje menihov, ki smo ga opisali, najustrezneje povzema razvrstitev s tremi skupinami:

$$C_1 = \{1, 2, 7, 12, 14, 15, 16\}$$

$$C_2 = \{4, 5, 6, 8, 9, 10, 11, 13\}$$

$$C_3 = \{3, 17, 18\}$$

Tedaj je glede na to razvrstitev preurejena matrika iz tabele 1 (permutirane vrstice in stolpci, tako da so najprej razvrščene enote prve skupine, nato druge in tretje) predstavljena v tabeli 2.

Tabela 2: Sampsonovi podatki – naklonjenost, permutirana matrika

		1	2	7	1	1	1	1						1	1	1		1	1
		1	2	7	2	4	5	6	4	5	6	8	9	0	1	3	3	7	8
John Bosco	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
Gregory	2	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mark	7	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Winfred	12	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hugh	14	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Boniface	15	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Albert	16	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Peter	4	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0
Bonaventur	5	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0
Berthold	6	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0
Victor	8	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0
Ambrose	9	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0
Romuald	10	0	0	0	0	0	0	0	1	1	0	0	1	0	0	1	0	0	0
Louis	11	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0
Amand	13	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Basil	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1
Elias	17	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
Simplicius	18	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0

Če v preurejeni matriki pregledamo strukturo povezav v posameznih pravokotnikih (blokih), opazimo, da so si menihi iz prve skupine kar močno naklonjeni in nihče od teh menihov ni naklonjen menihom iz druge in tretje skupine. Tudi menihi druge skupine so kar precej naklonjeni drug drugemu znotraj skupine in nenaklonjeni menihom iz drugih dveh skupin (z redkimi izjemami). Podobno velja tudi za menihe tretje skupine, ki so si med seboj naklonjeni, niso pa naklonjeni menihom iz drugih dveh skupin. Dobljeno strukturo lahko opišemo z bločno matriko

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

## Blok

Razvrstitev  $\mathcal{C}$  razbije relacijo  $R$  na *bloke*

$$R(C_i, C_j) = R \cap C_i \times C_j$$

Vsak blok sestavljajo enote, ki pripadajo skupinama  $C_i$  in  $C_j$ , in iz povezav, ki gredo iz skupine  $C_i$  v skupino  $C_j$ . Če je  $i = j$ , imenujemo  $R(C_i, C_i)$  *diagonalni blok*.

## **Bločni model**

Dana razvrstitev  $\mathcal{C}$  poraja *bločni model* takole:

vsako skupino razvrstitve stisnemo v točko modela; kateri šopi povezav med skupinami dajo povezave v modelu, pa je odvisno od izbrane vrste enakovrednosti.

Graf modela lahko predstavimo tudi z matriko, ki jo imenujemo *bločna matrika*.

## Enakovrednosti

### Strukturna enakovrednost

Enoti  $X$  in  $Y$  sta strukturno enakovredni, če je  $X$  povezan z vsako enoto iz množice  $E$  na enak način kot  $Y$ . Torej, če sta enoti  $X$  in  $Y$  strukturno enakovredni, sta zamenljivi (Lorrain in White, 1971).

Enoti  $X$  in  $Y$  sta *strukturno enakovredni*  $X \equiv Y$  natanko takrat, ko so izpolnjeni naslednji pogoji:

$$s1. \quad XRY \Leftrightarrow YRX$$

$$s2. \quad XRX \Leftrightarrow YRY$$

$$s3. \quad \forall Z \in E \setminus \{X, Y\} : (XRZ \Leftrightarrow YRZ)$$

$$s4. \quad \forall Z \in E \setminus \{X, Y\} : (ZRX \Leftrightarrow ZRY)$$

Iz definicije strukturne enakovrednosti izhaja, da so v tem primeru možne le štiri vrste idealnih blokov (Batagelj, Ferligoj in Doreian 1992).

0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	1
1	1	1	1	1	0	1	1	1
1	1	1	1	1	1	0	1	1
1	1	1	1	1	1	1	0	1
1	1	1	1	1	1	1	1	0

Slika 2: Idealni bloki za strukturno enakovrednost



## Regularna enakovrednost

Intuitivno, dve enoti sta regularno enakovredni, če sta enako povezani s skupinami enakovrednih enot (White in Reitz 1983).

Enakovrednost  $\approx$  na  $E$  je *regularna enakovrednost* na omrežju  $\mathcal{N} = (E, R)$  natanko takrat, ko za enote  $X, Y, Z \in E$ , iz  $X \approx Y$  izhajajo

$$\text{R1. } XRZ \Rightarrow \exists W \in E : (YRW \wedge W \approx Z)$$

$$\text{R2. } ZRX \Rightarrow \exists W \in E : (WRY \wedge W \approx Z)$$

Batagelj, Doreian in Ferligoj (1992) so pokazali, da regularni enakovrednosti ustrezata dve vrsti idealnih blokov:

- prazen blok (ki ima vse vrednosti 0)
- in regularni bloki (ki imajo v vsaki vrstici in vsakem stolpcu vsaj eno 1).

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	0	1	0	0
0	0	1	0	1
0	1	0	0	0
1	0	1	1	0

Slika 3: Idealna bloka za regularno enakovrednost

Vsaka strukturna enakovrednost je tudi regularna.

Strukturna enakovrednost je zelo stroga zahteva. Regularna enakovrednost je manj stroga in jo je mogoče pogosteje najti v omrežju.

## Določitev bločnih modelov

Problem bločnega modeliranja je poiskati tako razvrstitev enot omrežja, da se v razvrstitvi kar najbolje odraža izbrana vrsta enakovrednosti. V tem smislu je torej posebni problem **razvrščanja v skupine**, ki ga lahko zastavimo kot optimizacijski problem takole:

določi razrstitev  $\mathcal{C}^*$  iz množice vseh dopustnih razvrstitev  $\Phi$ , tako da je

$$P(\mathcal{C}^*) = \min_{\mathcal{C} \in \Phi} P(\mathcal{C})$$

kjer je  $P : \Phi \rightarrow \mathbb{R}$  *kriterijska funkcija*. Ta mora biti usklajena z izbrano vrsto enakovrednosti.

Kriterijsko funkcijo lahko sestavimo

- **posredno** kot funkcijo usklajene mere različnosti med pari enot z izbrano enakovrednostjo ali
- **neposredno** kot funkcijo, ki meri usklajenost razvrstitve s podatki o omrežju in izbrano enakovrednostjo.

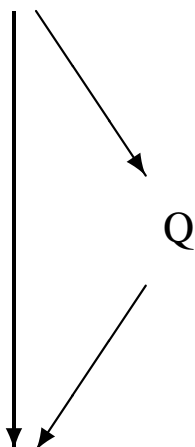
## Posredni pristop

RELACIJA

R

OPISI

ENOT



začetna relacija

matrika poti

triade

orbite

MATRIKA

RAZLIČNOSTI

D

OBIČAJNI

POSTOPKI

RAZVRŠČANJA

hierarhični postopki,

metoda prestavljanj, metoda voditeljev

Slika 4: Posredni pristop

Mera različnosti  $d$  je usklajena z izbrano enakovrednostjo  $\sim$ , če za vsak par enot velja

$$X_i \sim X_j \Leftrightarrow d(X_i, X_j) = 0$$

Za posredno določanje razvrstitev glede na strukturno enakovrednost se uporablja veliko različnih mer različnosti. Vendar jih je med njimi razmeroma malo zares usklajenih s to enakovrednostjo. Primer usklajene različnosti s strukturno enakovrednostjo je popravljena evklidska razdalja:

$$d(X_i, X_j) =$$

$$\sqrt{p * ((r_{ii} - r_{jj})^2 + (r_{ij} - r_{ji})^2) + \sum_{\substack{s=1 \\ s \neq i, j}}^n ((r_{is} - r_{js})^2 + (r_{si} - r_{sj})^2)}$$

$$p \in [0, 1, 2]$$

Posredni pristop ni primeren za določitev razvrstitev glede na regularno enakovrednost, ker niso znane mere različnosti, ki bi bile z njo usklajene.

## Postopek hierarhičnega združevanja v skupine

Najobsežnejši razred metod hierarhičnega razvrščanja v skupine predstavljajo metode, ki temeljijo na zaporednem združevanju (zlivanju) dveh ali več skupin v novo skupino. Med njimi so najpogostejše tiste, ki združijo vsakič po dve skupini. Te lahko v grobem opišemo z naslednjim postopkom:

vsaka enota je skupina:

$$C_i = \{X_i\}, X_i \in E, i = 1, 2, \dots, n$$

PONAVLJAJ, dokler ne ostane ena sama skupina:

določi najbližji si skupini  $C_p$  in  $C_q$ :

$$d(C_p, C_q) = \min_{u,v} d(C_u, C_v);$$

združi skupini  $C_p$  in  $C_q$  v skupino  $C_r = C_p \cup C_q$ ;

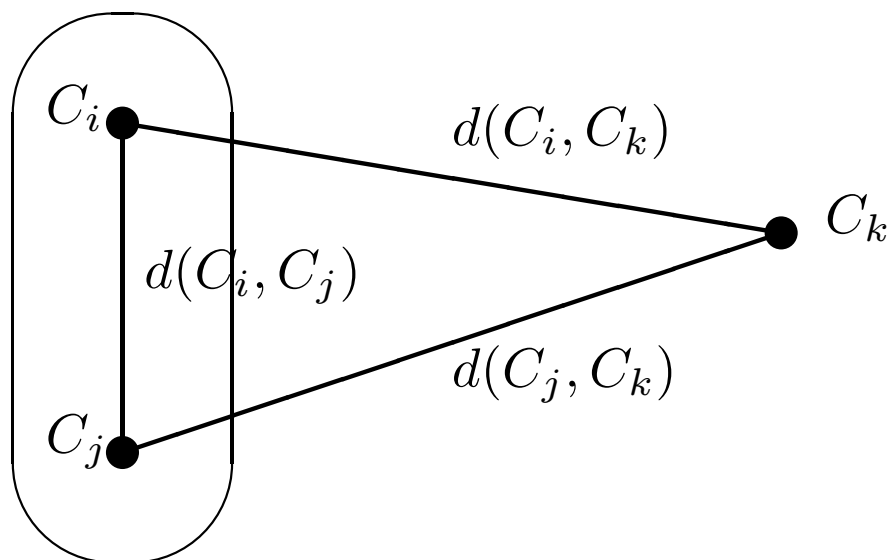
zamenjaj skupini  $C_p$  in  $C_q$  s skupino  $C_r$ ;

določi mere različnosti  $d$  med novo skupino  $C_r$  in ostalimi.

## Metode hierarhičnega združevanja

Mere različnosti  $d$  med novo skupino in ostalimi v postopku združevanja v skupine določamo na več načinov in ti določajo različne metode hierarhičnega združevanja v skupine.

Vzemimo, da imamo v nekem koraku postopka tri skupine  $C_i$ ,  $C_j$  in  $C_k$  ter podane mere različnosti med njimi približno tako, kot je prikazano na sliki.



Denimo, da sta skupini  $C_i$  in  $C_j$  najbližji, zato ju združimo v novo skupino  $C_i \cup C_j$ . Mero različnosti med novo skupino in skupino  $C_k$  lahko določimo na več načinov.

- **Minimalna metoda** ali tudi enojna povezanost (Florek et al. 1951; Sneath 1957):

$$d(C_i \cup C_j, C_k) = \min(d(C_i, C_k), d(C_j, C_k))$$

- **Maksimalna metoda** ali tudi polna povezanost (McQuitty 1960):

$$d(C_i \cup C_j, C_k) = \max(d(C_i, C_k), d(C_j, C_k))$$

- **McQuittyjeva metoda** (McQuitty 1966, 1967):

$$d(C_i \cup C_j, C_k) = \frac{d(C_i, C_k) + d(C_j, C_k)}{2}$$

- **Wardova metoda** (Ward 1963):

$$d(C_i \cup C_j, C_k) = \frac{(n_i + n_j)n_k}{(n_i + n_j + n_k)} d^2(T_{i,j}, T_k)$$

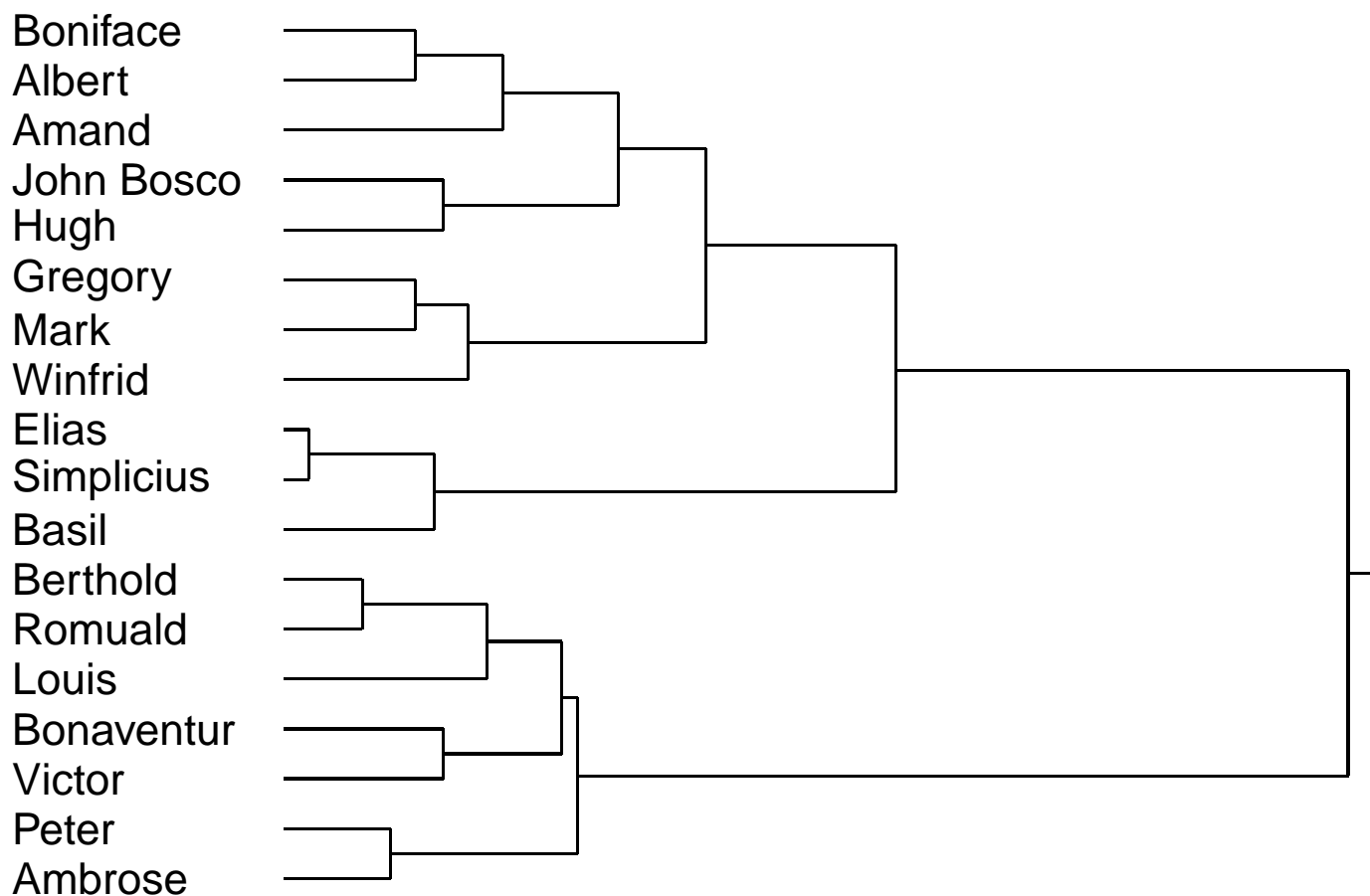
kjer s  $T_{ij}$  označimo težišče združene skupine  $C_i \cup C_j$  in s  $T_k$  težišče skupine  $C_k$ . Oznaka  $n_i$  pomeni število enot v skupini  $C_i$ .



## **Drevo združevanja**

Potek združevanja si lahko grafično ponazorimo z drevesom združevanja - dendrogramom. Listi tega drevesa so enote, točke združitve pa sestavljene skupine: levi in desni naslednik vsake točke sta skupini, iz katerih je nastala. Višina točke, ki jo imenujemo nivo združevanja, je sorazmerna meri različnosti med skupinama.

Primer drevesa združevanja je drevo menihov glede na njihovo naklonjenost v četrtem obdobju (glej relacijsko matriko v tabeli 1), kjer je bila upoštevana popravljena evklidska razdalja in Wardova metoda združevanja.



Slika 5: Drevo združevanja po Wardovi metodi

## Neposredni pristop

Kriterijsko funkcijo sestavimo tako, da meri odstopanje dejanskih blokov od pripadajočih idealnih blokov glede na model. (Batagelj, Doreian in Ferligoj, 1992; Batagelj, 1993; Doreian, Batagelj in Ferligoj, 1994).

Kriterijska funkcija  $P(\mathcal{C})$  mora biti **občutljiva** na obravano enakovrednost:

$P(\mathcal{C}) = 0 \Leftrightarrow$  s  $\mathcal{C}$  določeni bloki popolnoma ustrezajo izbrani enakovrednosti.

Izhajajmo iz razvrstitve  $\mathcal{C} = \{C_1, C_2, \dots, C_k\}$ , in označimo z  $\mathcal{B}(C_u, C_v)$  množico vseh idealnih blokov glede na blok  $R(C_u, C_v)$ . Potem lahko celotno napako razvrstitve  $\mathcal{C}$  izrazimo z naslednjo kriterijsko funkcijo

$$P(\mathcal{C}) = \sum_{C_u, C_v \in \mathcal{C}} \min_{B \in \mathcal{B}(C_u, C_v)} d(R(C_u, C_v), B)$$

kjer člen  $d(R(C_u, C_v), B)$  meri razliko (napako) med blokom  $R(C_u, C_v)$  in idealnim blokom  $B$ . Funkcija  $d$  mora biti usklajena z izbranim tipom enakovrednosti.

Tabela 3: Empirični bloki

	a	b	c	d	e	f	g
a	0	1	1	0	1	0	0
b	1	0	1	0	0	0	0
c	1	1	0	0	0	0	0
d	1	1	1	0	0	0	0
e	1	1	1	0	0	0	0
f	1	1	1	0	1	0	1
g	0	1	1	0	0	0	0

Tabela 4: Idealni bloki

	a	b	c	d	e	f	g
a	0	1	1	0	0	0	0
b	1	0	1	0	0	0	0
c	1	1	0	0	0	0	0
d	1	1	1	0	0	0	0
e	1	1	1	0	0	0	0
f	1	1	1	0	0	0	0
g	1	1	1	0	0	0	0

Tabela 5: Število napak - razlike med empiričnimi in idealnimi bloki

	A	B
A	0	1
B	1	2

Vrednost kriterijske funkcije je vsota vseh napak, v našem primeru je  $P = 4$ .

Za določitev iskane razvrstitve lahko uporabimo postopke lokalne optimizacije (Batagelj, Doreian in Ferligoj, 1992). Na primer metodo predstavljaj:

Določi začetno razvrstitev  $\mathcal{C}$ ;

**ponavljaj:**

če med tekočo razvrstitvijo  $\mathcal{C}$  in sosednjimi razvrstitvami obstaja razvrstitev  $\mathcal{C}'$ ,

za katero velja  $P(\mathcal{C}') < P(\mathcal{C})$

**potem** se pomakni v razvrstitev  $\mathcal{C}'$ , **sicer** končaj.

V tem postopku je *sosednost* razvrstitve določena:

- s *prestavitvijo* enote  $X_k$  iz skupine  $C_p$  v skupino  $C_q$ ;
- z *zamenjavo* enot  $X_u$  in  $X_v$  iz različnih skupin  $C_p$  in  $C_q$ .

Postopek lokalne optimizacije je dal naslednjo najboljšo razvrstitev Sampsonovih menihov glede na naklonjenost v četrtem obdobju:

Tabela 6: Sampsonovi podatki – naklonjenost

					1	1	1	1						1	1		1	1	1
		1	2	7	2	4	5	6	4	5	6	8	9	0	1	3	3	7	8
John Bosco	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0
Gregory	2	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mark	7	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Winfred	12	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hugh	14	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Boniface	15	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Albert	16	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Peter	4	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0
Bonaventur	5	0	0	0	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0
Berthold	6	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0
Victor	8	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0
Ambrose	9	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0
Romuald	10	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	0	0
Louis	11	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0
Basil	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
Amand	13	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
Elias	17	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Simplicius	18	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0