

**KOSTANTIN MOMIROVIĆ*, ANKICA HOŠEK,
KŠEŃIJA BOSNAR I FRANJO PROT**
Institut za kineziologiju Fakulteta za fizičku kulturu
u Zagrebu
Odjel za informatiku i statistiku

*Sveučilišni računski centar u Zagrebu
Odjel za informatičku matematiku i statistiku

ALGORITAM ZA DETEKCIJU KLIKA NA OSNOVI STRUKTURE KOMUNIKACIJSKIH MREŽA

Izvorni znanstveni članak
UDK 519.254:519.246.87:301.15
Primljeno 13. 10. 1983.

/ mikrosociologija / multi dimenzionalno skaliranje / komunikacijska mreža, analiza / spektralna dekompozicija /
/ binarna matrica /

Predložen je algoritam i napisan program za detekciju klika na osnovi strukture komunikacijske mreže definirane binarnom reprezentacijom komunikacijskih kanala između čvorova. Algoritam određuje klike parsimonijskim ortonormalnim transformacijama informatički značajnih lijevih i desnih svojstvenih vektora pridruženih nadprosječnim vrijednostima spektra mreže tako da jedna solucija određuje klike na temelju ulaznih, a druga na temelju izlaznih kanala. Pošanje algoritma prikazano je na podacima dobijenima sociometrijskim ispitivanjima jedne poluzatvorene grupe.

1. UVOD

Među uistinu impresivnom kolekcijom heurističkih postupaka za analizu sociometrijskih podataka predloženi su i neki čija je svrha detekcija klika (Katz, 1947; Festinger, Schackter i Beck, 1950; Coleman, 1964; Festinger, 1950; Luce i Perry, 1950; Nehnevajsa, 1974)² na temelju operacija nad sociometrijskim matricama. Najveći se broj tih postupaka svodi na potenciranje matrica, ili na prostu dekompoziciju tih matrica.

Detekcija klika je, naravno, taksonomski problem, i može biti rješavan efikasnim taksonomskim algoritmima. U ovom je radu predložen jedan takav algoritam koji određuje klike na temelju komunikacija između entiteta parsimonijskom transformacijom lijevih i desnih svojstvenih vektora binarne matrice koja opisuje strukturu komunikacijskih kanala.

2. ALGORITAM

Neka je $S = \{s_i, i=1, \dots, n\}$ grupa koju tvori n entiteta povezanih određenim brojem jednosmjernih ili dvosmjernih komunikacijskih kanala. Neka su intenzitet, frekvencija, opseg i sadržaj komunikacija između entiteta irelevantne varijable, tako da se struktura mreže koju tvore entiteti povezani komunikacijskim kanalima može svesti na binarni graf mreže.

¹ Izvrstan i vrlo sistematičan pregled tih postupaka vidi u I. Nehnevajsa, Soziometrie (In R. König, Grundlegende Methoden und Techniken der empirischen Sozialforschung, 10 Teil, Band, 2, pp. 268-285, Enke, Stuttgart, 1973).

² Opis ovih postupaka i popis referenci vidi u König, 1973.

Neka je $M = (m_{ij}); i=1, \dots, n; j=1, \dots, n$ matrica kojom je opisana struktura mreže i neka je ta matrica orijentirana tako da redci definiraju ulazne, a kolone izlazne kanale čvorova. Matrica M bit će općenito asimetrična, pozitivno semidefinitna matrica sa strukturom

$$M = X\Lambda X^{-1},$$

gdje je Λ dijagonalna matrica spektralnih vrijednosti, a X matrica njima pridruženih vektora.

Razmotrimo, međutim, matrice

$$R = M^T M$$

$$G = M M^T.$$

Kako je M pozitivno definitna ili pozitivno semidefinitna matrica, njena se struktura može napisati u obliku

$$M = Y\Lambda X^T,$$

gdje su Y i X matrice ortogonalnih svojstvenih vektora, normalnih tako da je $Y^T Y = I$ i $X^T X = I$. Otuda

$$R = X\Lambda^2 X^T$$

$$G = Y\Lambda^2 Y^T$$

(Green i Carrol, 1976; Fisher, 1978; Horst, 1965) pa se lijevi svojstveni vektori, pridruženi nenutlim svojstvenim vrijednostima matrice M mogu definirati kao

$$Y = M X \Lambda^{-1}$$

ako su vektori X definirani na matrici R^3 .

³ Naravno, osnova solucije može biti i matrica G , ili, jednostavno, spektralna dekompozicija matrice M .

Očito, u velikoj dijagonali matrice R je broj izlaznih kanala iz svakog čvora, a u vandijagonalnim članovima broj izlaznih kanala zajedničkih svakom paru čvorova. Naravno, u velikoj dijagonali matrice G je broj ulaznih kanala zajedničkih parovima čvorova.

Koordinatni sistem definiran lijevim svojstvenim vektorom Y matrice M određuje stoga položaj čvorova u prostoru ulaznih komunikacijskih kanala, a koordinatni sistem definiran desnim svojstvenim vektorima X matrice M određuje položaj čvorova u prostoru izlaznih komunikacijskih kanala.

Ovi su koordinatni sustavi, naravno, M — biortogonalni, jer

$$Y^T M X = X^T M^T Y = \Lambda.$$

Poredajmo svojstvene vrijednosti matrica R i G tako da je $\lambda_1^2 > \lambda_2^2 > \dots > \lambda_q^2$, gdje q posljednja nenulta svojstvena vrijednost, i organizirajmo tako i njima pridružene svojstvene vektore u matricama Y i X . Sada je matrica R i G moguće napisati kao zbroj matrica ranga jedan, dekomponiran u dvije komponente

$$R = \sum_{p=1}^k \lambda_p^2 X_p X_p^T + \sum_{p=k+1}^q \lambda_p^2 X_p X_p^T$$

odnosno

$$G = \sum_{p=1}^k \lambda_p^2 Y_p Y_p^T + \sum_{p=q+1}^q \lambda_p^2 Y_p Y_p^T.$$

Matrice

$$R^* = \sum_{p=1}^k \lambda_p^2 X_p X_p^T$$

odnosno

$$G^* = \sum_{p=1}^k \lambda_p^2 Y_p Y_p^T$$

su, za neki fiksni k , optimalna aproksimacija matrica R i G singularnim matricama R^* i G^* ranga $k < q < n$, pa je stoga i matrica

$$M^* = \sum_{p=1}^k \lambda_p^2 Y_p X_p^T$$

optimalna, za neki fiksni rang k , Eckart-Youngova aproksimacija matrice M (Eckart i Young, 1936).

Neka je $k < q$ izabran na neki pogodan način⁴, i neka je u matricama Y^* i X^* zadržano prvih k svojstvenih vektora, pridruženih zadržanim svojstvenim vrijednostima organiziranim u dijagonalnoj matrici Λ^* . Sada je

$$R^* = X^* \Lambda^2 X^{*T}$$

$$G^* = Y^* \Lambda^2 Y^{*T}$$

i

$$M^* = Y^* \Lambda^* X^{*T}.$$

Ako čvorovi tvore klike, tj. nakupine čvorova povezanih velikim brojem komunikacijskih kanala, sa slabim vezama s ostalim čvorovima ili drugim klikama, te je klike moguće detektirati nekom parsimonijskom transformacijom matrica Y^* i X^* .

Neka su Q i T ortonormalne transformacijske matrice određene tako da elementi matrica

$$P = Y^* Q = (p_{ip}) \quad \begin{matrix} i = 1, \dots, n \\ p = 1, \dots, k \end{matrix}$$

i

$$A = X^* T = (a_{ip}) \quad \begin{matrix} i = 1, \dots, n \\ p = 1, \dots, k \end{matrix}$$

ekstremiziraju neku parsimonijsku funkciju.⁵

Ako klike postoje, i ako je k određen sukladno broju klika, u vektorima p_p , $p=1, \dots, k$ matrice P imat će elemente značajno različite od nule čvorovi koji tvore klike na temelju ulaznih kanala, a u vektorima A_p , $p=1, \dots, k$ matrice A imat će elemente značajno različite od nule čvorovi koji tvore klike na temelju izlaznih kanala.

Vektori u matricama P i A i dalje su, naravno, ortogonalni. Međutim, ta dva sistema više nisu biortogonalna, jer

$$P^T A = Q^T \Lambda^* T.$$

što u općem slučaju nije dijagonalna matrica.

Ako projiciramo vektore ulaznih kanala u sistem definiran grozdovima izlaznih kanala, i ako projiciramo vektore izlaznih kanala u sustav definiran grozdovima ulaznih kanala, tj. ako izvedemo operacije

$$K = M X^* T$$

i

$$L = M^T Y^* Q.$$

možemo utvrditi relacije grozdova ulaznih kanala operacijom

$$L^T L = Q^T \Lambda^2 Q = W$$

i relacije grozdova izlaznih kanala operacijom

$$K^T K = T^T \Lambda^2 T = V.$$

Neka je $D_W = (\text{diag} W)^{-1/2}$ i neka je $D_V = (\text{diag} V)^{-1/2}$.

Sada su kosinusi kuteva vektora grozdova ulaznih kanala u prostoru izlaznih kanala elementi matrice

$$C_L = D_W W D_W$$

⁴ U programu ITA primijenjena je jednostavna Guttman-Kaiserova strategija:

$$k = \text{num} \left(\lambda^2 > \sum_{p=1}^q \lambda_p^2 / q \right)$$

⁵ U programu ITA ekstremizirana je Kaiserova (1958) varimax funkcija

$$W = \sum_{i=p}^n \sum_{p=1}^k e_{ip}^4 \quad - \sum_{p=1}^k \left(\sum_{i=1}^n e_{ip}^2 \right)^2$$

za $e_{ip} = p_{ip}$, a_{ip} .

a kosinusi kuteva vektora grozdova izlaznih kanala u prostoru ulaznih kanala elementi matrice

$$C_k = D_v V D_v.$$

Ova operacija omogućuje i određivanje kosinusa kuteva vektora grozdova ulaznih i izlaznih kanala, jer je

$$L^T K = Q^T A^3 T,$$

a kosinusi kuteva vektora grozdova definiranih na temelju ulaznih i izlaznih kanala elementi matrice

$$C_{Lk} = D_w Q^T A^3 T D_v.$$

3. PROGRAM ITA

Program ITA napisan je u verziji 5 SS jezika i implementiran u programskim bibliotekama SRCE*SS-MAKRO⁶ i FFK*LIB⁷. Program dopušta grupe do 250 entiteta i pretpostavlja da su u recima matrice podataka vektori ulaznih, a u kolonama te matrice vektori izlaznih kanala. U zapisu koji sadrži podatke imenima entiteta valja pridružiti oznaku da se radi o ulaznim kanalima. Izlazne kanale program tretira kao varijable, pa imena varijabli moraju biti identična imenima entiteta kojima je pridružena oznaka da se radi o izlaznim kanalima.

4. REZULTATI ANALIZE JEDNE MALE POLUZATVORENE GRUPE

Analizirana je struktura jedne male grupe čija je djelatnost u velikoj mjeri ovisila o kooperaciji članova grupe.⁸ Entiteti su tretirani kao čvorovi mreže, a komunikacijski kanal određenog smjera smatran je otvorenim, ako je intenzitet komunikacija između ma koja dva čvora prešao neku liminalnu vrijednost.

U tabeli 1 dati su rezultati dobijeni analizom strukture ulaznih, a u tabeli 2 rezultati dobijeni analizom strukture izlaznih kanala.

Tabela 1

FORMACIJA KLIKA NA OSNOVU ULAZNIH KANALA

	U VRX 1	U VRX 2	U VRX 3
U 07	-.0306	(.3699)	-.1245
U 08	.0687	(.5138)	-.0362
U 09	-.1498	(.3089)	.1957
U 10	.0142	-.0568	(.4707)
U 11	-.0646	.1734	(.5500)
U 12	.0125	-.0552	(.4623)
U 13	(.4968)	-.1731	.2527
U 14	(.5607)	.0589	-.2463
U 17	(.5976)	.0762	.0900
U 19	.0017	(.4440)	.1845
U 20	-.0306	(.3699)	-.1245
U 21	.0798	-.0895	(.1545)
U 02	.2045	(.2937)	-.0324

Tabela 2

FORMACIJA KLIKA NA OSNOVU IZLAZNIH KANALA

	K VRX 1	K VRX 2	K VRX 3
K2 07	-.0845	(.5171)	-.0438
K2 08	.0443	(.5080)	.1057
K2 09	(.3535)	.2950	-.1660
K2 10	(.3884)	.2941	-.1583
K2 11	(.4786)	-.1504	.0441
K2 12	(.3797)	-.0915	.2842
K2 13	(.5617)	-.0764	-.0036
K2 14	-.0472	.1483	(.3414)
K2 17	-.0279	.0196	(.4611)
K2 19	.0126	.1285	(.4070)
K2 20	-.0945	(.3222)	.1202
K2 21	.1168	-.1405	(.2801)
K2 02	-.0247	-.0467	(.5137)

Tri klike formirane na osnovu ulaznih kanala u ne-negativnim su međusobnim vezama (tabela 3); no te su veze niže od osrednjih. Prvu kliku tvore entiteti koji su u grupu došli relativno kasno i ne potiču iz sredine iz koje potiču ostali članovi grupe. Druga, po broju najveća klika sastavljena je od entiteta superiorne profesionalne efikasnosti. Treću kliku tvore tri entiteta sa vrlo visokim nivoom aspiracije i naglim trendom postizanja vrhunske profesionalne efikasnosti. Posljednja dva entiteta od kojih prvi pripada trećoj, a drugi drugoj kliku zapravo su bili autsajderi pod vidom njihove profesionalne efikasnosti u vrijeme kada je provedeno ovo istraživanje.

Klike formirane na temelju izlaznih kanala suštinski se razlikuju od klika formiranih na temelju ulaznih kanala. Relacije između ovako formiranih klika su također nenegativne, ali i dalje jedva osrednje (tabela 4). Kriterij na osnovi kojeg su se formirale klike bio je različit od kriterija koji je formirao klike na osnovi ulaznih kanala. Prvu kliku tvorila je relativno homogena grupa entiteta suprotstavljenih politički rukovodstva grupe. Drugu, slabu kliku sačinjavalo je nekoliko entiteta s izrazitom sklonošću ka kooperaciji. Treću kliku sačinjavao je ostatak grupe, koji se prema rukovodstvu odnosio neutralno.

Tabela 3

RELACIJE KLIKA ODREĐENIH NA OSNOVI ULAZNIH KANALA

	U VRX 1	U VRX 2	U VRX 3
U VRX 1	1.0000	.5313	.5789
U VRX 2	.5313	1.0000	.3741
U VRX 3	.5789	.3741	1.0000

⁶ Javna biblioteka Sveučilišnog računskog centra u Zagrebu.

⁷ Zatvorena biblioteka Računskog centra Fakulteta za fizičku kulturu u Zagrebu.

⁸ Priroda djelatnosti grupe bila je takva da su članovi grupe najveći dio ne samo radnog, već i slobodnog vremena morali provoditi zajedno.

Tabela 4

RELACIJE CLIKA ODREĐENIH NA OSNOVI IZLAZNIH KANALA

	K VRX 1	K VRX 2	K VRX 3
K VRX 1	1.0000	.4387	.5923
K VRX 2	.4387	1.0000	.4822
K VRX 3	.5923	.4822	1.0000

Profesionalna efikasnost grupe u vrijeme kada je provedena ova analiza bila je vrlo visoka, a standardne sociometrijske procedure govore su u prilog hipotezi da je kolektiv vrlo homogen. Međutim, samo godinu dana kasnije došlo je do otvorenog sukoba između prve klike formirane na osnovu izlaznih kanala i rukovodstva grupe. Dio članova grupe napustio je kolektiv, a rukovodstvo grupe bilo je uklonjeno. Heterogenost kolektiva postala je sasvim očita, a profesionalna efikasnost je drastično opala.

Čini se, prema tome, da ovaj algoritam može biti od neke koristi za detekciju klika, posebno u slučajevima kada postojanje tih klika ne može biti utvrđeno jednostavnom inspekcijom sociometrijskih matrica ili primjenom uobičajenih procedura za detekciju klika, koje se temelje na potenciranju sociometrijskih matrica.

LITERATURA

1. Eckart, C. and G. Young: The approximation of one matrix by another of lower rank. *Psychometrika*, 1 (1936), 211—218.
2. Fishera, G.: *Numerical and quantitative analysis*. Pitman, London, 1978.
3. Green, P. E. and J. D. Carroll.: *Mathematical tools for applied multivariate analysis*. Academic Press, New York, 1976.
4. Horst, P.: *Factor analysis of data matrices*. Holt, Rinehart and Winston, New York, 1965.
5. Kaiser, H. F.: The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23 (1958), 187—200.
6. König, R.: *Grundlegende Methoden und Techniken der empirischen Sozialforschung*. 1^o Teil, aBnd 2, 3^o Auflage. Enke Verlag, Stuttgart, 1973.

Momirović, K.; Hošek, A.; Bosnar, K.; Prot, F.:

THE ALGORITHM FOR DETECTION OF CLIQUES ON THE BASIS OF STRUCTURE OF COMMUNICATION NETWORKS

microsociology / multidimensional scaling / communication network analysis / spectral decomposition of binary matrices

The algorithm is proposed and the program written for detection of cliques on the basis of structure of a communication network defined by the binary representation of communication channels between knots.

The algorithm determines the cliques with the parsimonious orthonormal transformations of informatically significant left and right characteristic vectors associated above-average values of the spectrum of network, in such a way that one solution determines the cliques on the basis of the input channels and the other on the basis of the output channels.

The behaviour of the algorithm is shown on the data obtained in sociometric studies of a semi-closed group.

Константин Момирович, Анкица Хошек, Ксения Боснар и Франье Прот

АЛГОРИТМ ДЛЯ ОБНАРУЖЕНИЯ КЛИК НА ОСНОВЕ СТРУКТУРЫ КОММУНИКАЦИОННЫХ СЕТЕЙ

Предложен алгоритм и написана программа для обнаружения клик на основе структуры коммуникационной сети, представленной двойными коммуникационными каналами между узлами. Алгоритм определяет клики при помощи парсимонических ортогональных трансформаций значительных с информационной точки зрения, левых и правых характерных векторов, которые присоединяются к сверхнормальным величинам спектра сети, так что при помощи одного решения клики определяются на основе входных, а при помощи другого — на основе выходных каналов. Принципы работы алгоритма показаны на данных социометрического исследования одной полужакрытой группы.