

EVALUAREA PARAMETRILOR UNOR REȚELE COMPLEXE DIN DOMENIUL FINANCIAR – BANCAR

Radu-George Anghel

e-mail: Anghel.Radu@gmail.com

Mihai Tertîșco

e-mail: mihai_tertisco@yahoo.com

Universitatea Politehnica București

Rezumat: În lucrare se propun variante matricial-vectoriale pentru calculul unor indicatori ai distribuției probabilității gradelor și a măsurii nivelului de apropiere a nodurilor rețelelor complexe din domeniul financiar-bancar, în vederea depistării în rețea a unor aglomerări de arce și noduri numite COMUNITAȚI. În acest scop, s-a introdus noțiunea de vector caracteristic al unui nod din rețea. Acest vector este singura informație inițială, disponibilă pentru evaluarea parametrilor rețelei. Totalitatea vectorilor caracteristici formează matricea de adiacență a rețelei analizate. Se are în vedere numai tipul rețelelor reprezentabile prin grafuri neorientate. Sunt dezvoltate și adaptate, ca indicatori de evaluare a prezenței aglomerărilor în rețea, unele concepte cunoscute, cum ar fi: abaterea pătratică medie dintre vectorii caracteristici ai nodurilor, distanța dintre vectorii caracteristici, corelația dintre acești vectori etc. Este prezentat un studiu de caz.

Cuvinte cheie: rețea complexă, distribuția probabilității, coeficient de corelație, vectori proprii ai matricei de incidență, distanța.

1. Elemente preliminare privind rețelele complexe

Într-un sistem social, mulți agenți economici colaborează, interacționând unii cu ceilalți și influențându-se reciproc. Aceste colaborări au loc în cadrul unor grupuri de colaboratori preferați, în funcție de specificul activității, sau grupați după alte criterii ori interese. Studiul și analiza unor astfel de sisteme complexe reale se poate face pe baza unor date experimentale, stocate în baze de date despre legăturile și tranzacțiile financiar-bancare dintre agenții care fac parte din marea familie a agenților comerciali [5], [6].

Fie mulțimea firmelor $V = \{v_1, v_2, v_3, v_4, \dots, v_N\} = \{1, 2, \dots, N\}$, reprezentate prin aceeași bancă financiară. Legăturile dintre aceste firme sunt exprimate de tranzacțiile financiare, care sunt reprezentate prin transferurile dintr-un cont în altul. Dacă între oricare pereche de conturi există câte o singură tranzacție, numărul total de tranzacții posibile este:

$$M = N(N-1)/2 \quad (1)$$

și corespunde numărului total de legături posibile între diverse perechi de noduri din cele N . Spre exemplu, dacă $N=3$ cele trei legături posibile sunt: (1,2), (1,3) și (2,3). Dacă însă se consideră legături posibile ale nodurilor cu ele înșile (bucle elementare în rețea), atunci numărul total de legături posibile este:

$$m = n(n+1)/2 \quad (2)$$

în care n reprezintă numărul de noduri al rețelei, iar m este numărul total de arce între fiecare și toate celelalte inclusiv arcele de la fiecare nod la el însuși. Pentru a trage o concluzie referitoare la complexitate, subliniem că există întotdeauna și alte procedee de modelare a sistemelor complexe, decât cele clasice, bazate pe divizarea acestora în componente simple (primare). Unul dintre aceste noi procedee ar putea consta în simularea întregului sistem cu toate componentele sale pentru a identifica și evalua numeric anumite proprietăți ale sistemului în ansamblu. Stabilirea unor proprietăți ale grafurilor ajută la studierea rețelelor pe care le reprezintă. Evaluarea incertă (fuzzy) a unora din proprietăți este, de cele mai multe ori, preferabilă față de rezolvarea precisă, matematică, deoarece în cazul rețelelor cu grad ridicat de complexitate este mult mai importantă evoluția și proprietățile rețelei, decât rezultatul matematic ce decurge din analiza grafului rețelei.

„Sistemele complexe” nu sunt încă foarte bine definite în domeniul științific care le studiază și nici ca limbaj și terminologie nu s-au stabilizat. O definiție uzuală ar fi: *un sistem complex este un sistem ce conține mai multe entități care interacționează, entități al căror comportament nu poate fi explicat singular astfel ca să se evalueze comportamentul sistemului în ansamblu, pornind de la comportamentul unei entități individuale.*

Clasificarea rețelelor complexe se face după diverse criterii, cum ar fi:

- specificul domeniului real din care provine procesul pe care-l reprezintă (rețele economice; rețele tehnologice; rețele informatice; rețele de suprastructură; rețele hidrografice; rețele biologice; rețele de învățământ; rețele sociale; rețele de colaborare științifică etc.);
- tipul de graf al rețelei (graf neorientat-figura 2, graf orientat, graf neorientat bipartit, graf orientat bipartit-rețea Petri, rețele mixate, rețele comunitare (figura 1) etc.);

- caracterul determinist ori aleator al sistemului real (rețele deterministe, rețele stocastice);
- extinderea geografică (rețele globale, rețele naționale, rețele regionale, rețele locale);
- numărul egal de vecini ai tuturor nodurilor (rețele regularizate-de tip grilă cu ochiuri egale, rețele neregularizate).

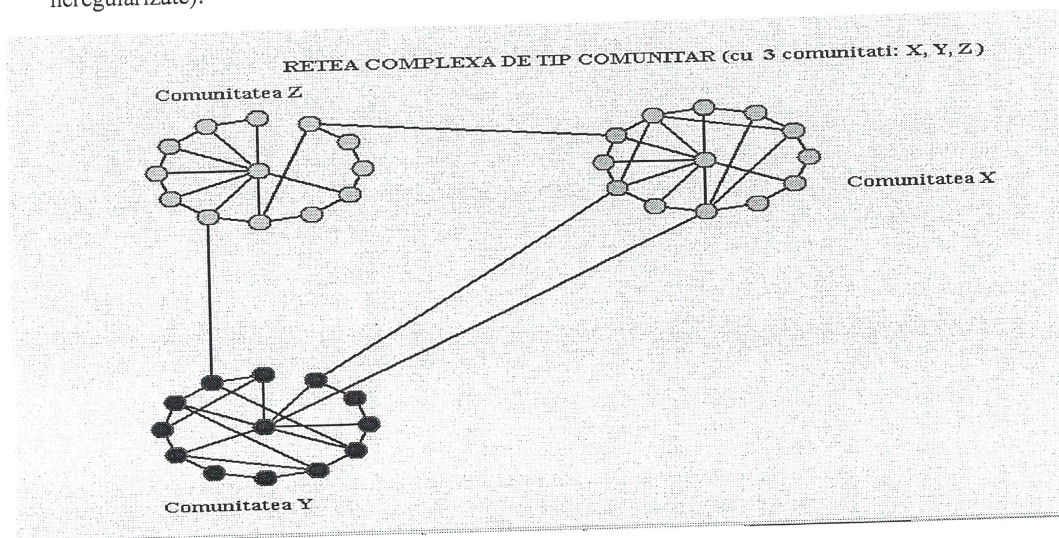


Figura 1. O rețea de legături între agenți economici de două categorii (IMM-uri diverse, grupate în jurul unor întreprinderi mari, care se află în centrul diverselor aglomerări sau clusteri)

Pentru 1000 de firme reprezentate de către o bancă financiară rezultă un număr de peste jumătate de milion de tranzacții. Dacă se ține cont de faptul că există mai multe tranzacții între firmele din aceeași pereche rezultă dimensiuni enorme ale băncii de date, care păstrează istoricul tranzacțiilor financiare. Aceste date stocate sunt utilizate pentru studiul rețelei respective. O primă utilizare a acestor date existente este construirea unei imagini grafice, precum cea din figura 1, în care mulțimea nodurilor grafului reprezintă totalitatea agenților, iar arcele-liniile de legătură dintre noduri-exprimă tranzacții financiare dintre diverși agenți. În acest graf, mulțimea nodurilor este $V = \{V_1, V_2, \dots, V_N\} = \{1, 2, 3, \dots, N\}$, iar mulțimea tranzacțiilor (arcelor), între conturile celor N firme, este $E = \{e_1, e_2, \dots, e_m\}$.

Acestui graf al rețelei complexe i se pot pune în corespondență cei N vectori de vecini ai tuturor nodurilor. Acești vectori caracteristici ai nodurilor rețelei sunt utilizați pentru calculul unor indicatori de caracterizare locală, care permit să se răspundă la întrebări de genul:

- Care este lungimea medie a unui lanț (drum) de firme colaboratoare?
- Care este numărul de colaboratori direcți ai fiecărei firme?
- Care este densitatea legăturilor în rețea a fiecărui agent în raport cu ceilalți?
- etc.

Un al doilea scop al utilizării datelor empirice, care au stat la baza construirii grafului, este obținerea unor răspunsuri privitoare la comunitatea agenților economici, reprezentați prin nodurile rețelei complexe, construită pe baza datelor empirice inițiale și a caracteristicilor locale, obținute la pasul precedent. Prin analiza statistică a rețelei se caută răspunsuri la întrebări de genul:

- care sunt grupurile de agenți uniți între ei prin legături de colaborare și interese reciproc avantajoase?
- cât de intense sunt legăturile dintre aceste grupuri la care liantul dintre agenți este colaborarea cu firma fanion din grup?
- care din agenți au influență mai mare asupra altora?
- există fărâmițări (descompuneri în grupuri mici) și ce a determinat această fărâmițare?
- care din conexiuni sunt cruciale pentru funcționarea grupului?

A treia destinație a datelor empirice, după prelucrarea lor statistică, este construirea unui model matematic sau a unui model computerizat imitativ (metoda Monte Carlo) pentru procesele care au loc în sistemul rețelizat. Acest model permite:

- efectuarea unor analize privind prognoza evoluției în viitor a rețelei;
- prevenirea unor catastrofe;
- apărarea rețelei de anumite atacuri din exteriorul rețelei, în cadrul luptei concurențiale legale dintre bănci;
- depistarea scurgerilor de informații din rețea etc.

Aceste probleme specifice rețelelor complexe sunt rezolvate pe baza informațiilor extrase din matricea de adiacență și unele caracteristici locale ale nodurilor rețelei.

Un vector A_i asociat nodului i din graf are N elemente ($k=a_{i1} \dots a_{ik} \dots a_{iN}$) corespunzătoare celor N noduri ale rețelei:

$$A_i = [a_{i1}, a_{i2}, \dots, a_{iN}], \quad (3)$$

în care $a_{ij}=1$ dacă nodul j din rețea este vecinul lui i (adică, dacă între nodul i și nodul j există un singur arc de lungime $L=1$) și $a_{ij}=0$ în caz contrar. Vectorul (3) îl numim vector de vecinătate sau vector caracteristic al nodului i .

O caracteristică importantă a fiecărui nod este gradul nodului sau, respectiv, numărul de vecini ai nodului. Acesta este calculat prin însumarea termenilor vectorului A_i . Pentru nodul i , gradul este notat k_i :

$$k_i = a_{i1} + \dots + a_{iN} = A_i \times A_i', \quad (4)$$

în care s-a notat cu A_i' vectorul caracteristic transpus.

Un alt concept în teoria grafurilor este drumul dintre două noduri i și j . Acesta este definit ca o secvență de noduri consecutive $\{i+1, i+2, \dots, j-1, j\}$ sau, respectiv, secvența de arce care leagă nodul i de nodul j : $\{(i+1, i+2), (i+2, i+3), \dots, (j-1, j)\}$. Un circuit este un drum care se închide în nodul din care a pornit. Un circuit de 3 arce se numește triunghi. Un drum elementar este un drum în care nici un nod nu apare de mai multe ori. Un circuit elementar este un drum în care nu apare decât primul nod de două ori. Lungimea unui drum este numărul de arce pe care le conține sau numărul de noduri minus 1. Drumul, pornind din nodul i și terminându-se în nodul j , cu cea mai mică lungime posibilă, se numește geodezică dintre i și j . Lungimea geodezice dintre i și j se numește distanța dintre i și j , se notează cu $d(i, j)$ și exprimă numărul de arce, pe cel mai scurt drum dintre i și j . Mulțimea de noduri (T_i) plasate la distanța 1 de nodul i se numește vecinătatea lui i . Distanța maximă dintre i la oricare alt nod din graful G se numește centricitatea nodului i . Cea mai mare valoare a centricității dintre toate nodurile se numește diametrul grafului, iar valoarea minimală se numește raza grafului. Distanței medii i se spune des caracteristica lungimii drumurilor. Acești parametri se pot stabili pe baza vectorilor de vecinătate ai nodurilor. O submulțime W de noduri cu o submulțime de arce (definite pe W cu reprezentare în E) se numește subgraf. Se spune că un subgraf este maximal în raport cu anumite proprietăți dacă nu poate fi extins fără pierderea proprietății respective. Dacă vorbim despre subgraful maximal al unei mulțimi de noduri $V' \subseteq V$, ne referim la graful (V', E') , unde E' este o submulțime $\{i, \dots, j\} \in E$, unde $i, j \in V$. Un subgraf maximal dintre două noduri se numește diadă; un subgraf maximal de trei noduri și două arce, se numește triadă, un subgraf maximal de trei noduri și trei arce, pe care se poate executa o deplasare în circuit închis, parcurgând toate cele trei arce, se numește triunghi.

Cercetările experimentale, [1], au consemnat faptul că numeroase rețele din domeniul financiar-bancar au un aspect de structură comunitară ("community structure") adică, grupuri relativ distincte de noduri și arce. Agenții economici au tendința să dezvolte afaceri și tranzacții financiar-bancare în cadrul unor grupuri (comunități) la care se afiliază, conform unor interese similare, domenii comune de activități și afaceri sau a unor afecțiuni reciproce. În modelul adoptat al rețelei, fiecare din agenți este reprezentat printr-un nod, iar legăturile tranzacționale dintre ei, prin arce. Fiecare din grupuri formează așa - zise comunități de noduri legate între ele printr-o mare densitate de arce, iar comunitățile între ele sunt conectate printr-o densitate mică de arce, ca în figura 1, în care sunt evidente trei astfel de comunități: X , Y și Z .

Scopul acestei lucrări este elaborarea și testarea unor metode destinate identificării aglomerărilor de noduri de tip comunitar pe baza evaluării similitudinii nodurilor perechi. Cunoașterea caracteristicilor unei rețele permite o alocare mai rațională a resurselor și o exploatare eficientă a acesteia, în concordanță cu interesele clienților, dar și cu scopul asigurării unui randament adecvat funcționării acesteia. În cazul unor rețele de mari dimensiuni, nici nu poate fi vorba de rezolvarea acestor aspecte doar empiric, prin inspectarea vizuală a datelor disponibile privind tranzacțiile financiar - bancare dintre clienți. Pentru analiza rețelei asistată de calculator și identificarea grupurilor comunitare, se întrevăd trei căi.

Prima cale constă în selectarea unor subgrafuri corespunzătoare în rețea pe baza identificării unor noduri sau a unor arce cu caracteristici speciale, prezente în rețea.

Două noduri din rețea sunt considerate structural echivalente dacă ambele au aceiași vecini. Echivalența structurală exactă este rar întâlnită, dar echivalența structurală aproximativă este folosită ca bază, în așa-numita metodă a clusterizării ierarhizate, descrisă mai jos. De exemplu, dacă se constată că cinci noduri au mulți vecini comuni, aceasta poate fi un semn că avem de-a face cu un grup comunitar. Această metodă de identificare în rețea a unor aglomerări de arce de tip comunitar o numim „metoda analizei arcelor”.

A doua cale constă în determinarea vectorilor proprii ai matricei de incidență a rețelei și pe baza unor proprietăți ale acestor vectori se identifică aglomerările de tip comunitar, prezente în rețea. Structura comunitară a rețelei poate fi evidențiată și prin inspectarea matricei de adiacență, după cum se poate constata din exemplul de rețea din figura 2 (în care este prezentat atât graful, cât și matricea de adiacență a rețelei). Întrucât rețelele complexe din domeniul financiar – bancar sunt de foarte mari dimensiuni, metoda depistării aglomerarilor de arce, prin inspectare vizuală a grafului rețelei sau a matricei de adiacență, este exclusă. Identificarea acestor aglomerări se poate face numai prin metode de calcul adecvate cu asistarea de către calculator a întregului proces de identificare.

2. Caracteristici structurale cantitative ale nodurilor și arcelor [4]

Dacă prima secțiune furnizează informații despre modul în care se poate evalua probabilitatea existenței unor clienți valoroși ai băncii, în secțiunea 2, se determină indicatori ai rețelei care permit să se precizeze care sunt acești clienți valoroși ai băncii.

2.1. Gradul și distribuția gradelor în rețelele complexe. Studiu de caz

Gradul poate constitui un indicator deosebit de util pentru analiza schemelor complexe, asociate rețelelor. Gradul este un mod de a măsura importanța unui nod și chiar a unui grup. Pentru multe tipuri de rețele reale, nodurile, caracterizate prin grade de nivel ridicat, au rolul de centru pentru mulțimea înconjurătoare de noduri de grad mai scăzut. Se poate spune că gradul este o măsură a centralității locale. Principala caracteristică de structură a unei rețele este distribuția de densitate a gradului – distribuția graduală. Datele experimentale au condus la următoarea constatare privind distribuția gradelor în cazul rețelelor complexe din domeniul financiar bancar: „în rețea sunt dese nodurile de grad scăzut și sunt rare cele de grad înalt”. Această constatare s-a făcut pe o rețea cu $N=1000$ de noduri care corespund celor 1000 de agenți economici care efectuează tranzacții în cadrul aceiași bănci financiare. Legăturilor tranzacționale dintre clienții băncii, le corespund arcele între nodurile rețelei analizate. În cadrul analizei, se efectuează o ierarhizare a nodurilor după frecvența de apariție a diferitelor grade. Spre exemplu, în studiul de caz făcut, s-a constatat că, din cele 1000 de noduri, 175 au un singur vecin, deci gradul $k=1$, iar frecvența de apariție este $p(k=1)=175/1000$. Pentru nodurile de grad $k=2$, s-a observat o frecvență de apariție mai mică $p(k=2)=90/1000$ etc. Dacă N este suficient de mare, o astfel de frecvență, $p(k)$, exprimă probabilitatea de a întâlni un nod de grad k , în cazul inspectării nodurilor din rețea.

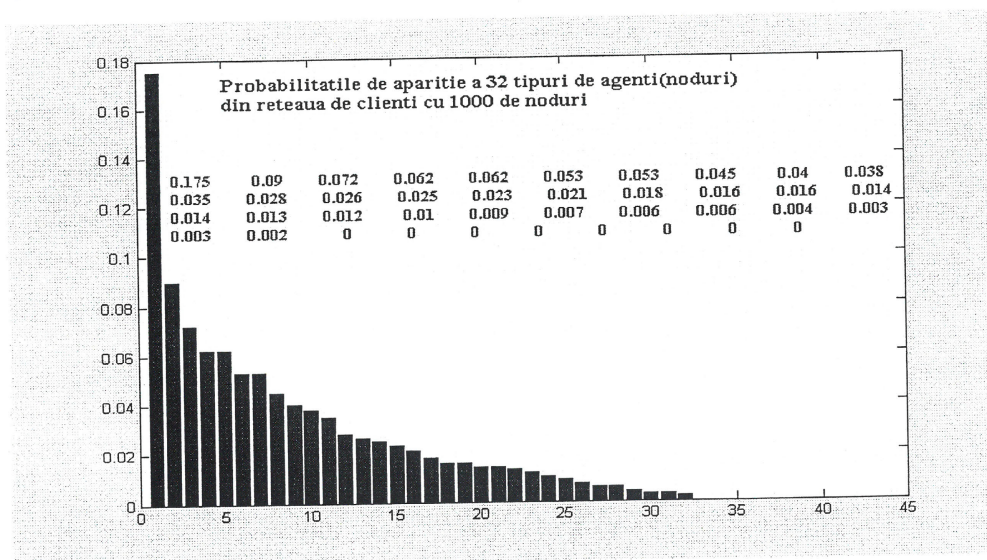


Figura 2. Histograma unei rețele complexe din domeniul financiar - bancar, cu 1000 de noduri clasificate în 32 tipuri de noduri care diferă între ele prin numărul de vecini

Acest tip de distribuție specifică unor rețele din lumea reală este denumită distribuție exponențială (adică funcția de densitate a probabilității $p(k)$ este o funcție exponențială de gradul k al nodurilor). Această funcție de distribuție a probabilităților, de a întâlni în rețea un nod de grad k , este de forma:

$$p(k) = ak^{-b}, \quad (5)$$

unde a și b sunt constante [4].

Importanța cunoașterii funcției (5) de distribuție a probabilităților gradelor în rețeaua tranzacțiilor bancare, derivă din faptul că, pe baza acesteia, se pot răspunde la întrebări de genul:

- care este probabilitatea ca în bancă să existe clienți care efectuează tranzacții financiare lunare cu cel puțin 10 clienți ai băncii ?, dar cu 40?;
- cunoscând parametrii funcțiilor de distribuție, $P_1(k)$ și $P_2(k)$, pentru două societăți financiare diferite, se poate răspunde la întrebarea, „care dintre bănci are mai mulți clienți valoroși?” etc.

2.2. Estimarea parametrilor distribuției exponențiale

Parametrii distribuției exponențiale (5) pot fi estimați pe baza datelor experimentale (de exemplu, cele afișate în figura 2). În continuare, au fost propuse și testate trei metode de estimare a parametrilor din modelul propus pentru funcția de distribuție a probabilităților de apariție în rețea a unui nod de grad k . Metoda cea mai adecvată pentru estimare în studiul de caz va fi considerată cea pentru care suma pătratelor erorilor este cea mai mică.

Metoda 1 de estimare. Parametrul a poate fi aproximat prin valoarea probabilității nodului de grad unitar: $a=p(k=1)$, în acest caz, iar relația (5) devine:

$$P_1(k) = p(k=1) k^{-b_1}, \quad (6)$$

în care $p(k=1)$ este determinat din datele experimentale. În studiul de caz, pentru care rezultatele sunt afișate în figura 2, rezultă $p(k=1)=0,175$. Celălalt parametru, b_1 , al funcției exponențiale de distribuție, rezultă din (6), pentru $k=10$:

$$b_1 = \log p(k=1) - \log(p(10))$$

Pentru datele experimentale, din studiul de caz, se obține,

$$b_1 = \log(.175) - \log(.0038) = 1.663,$$

iar modelul din (6) devine:

$$p_1(k) = 0,175 k^{-1.66} \quad (7)$$

Erorile $e(k)$, de reprezentare a datelor experimentale $p(k)$ prin modelul $p_1(k)$ din (7), sunt exprimate prin diferențele, $e(k) = p(k) - p_1(k)$.

Suma pătratelor erorilor SPE1, în acest caz devine:

$$SPE1 = \sum_{k=1}^{40} (p(k) - 0,175 \cdot k^{-1.66})^2 = 0,02 \quad (8)$$

în care $p(k)$ reprezintă cele 40 de valori numerice ale probabilităților, determinate experimental și afișate în figura 2.

Metoda 2 de estimare. În acest caz, se consideră cunoscut parametrul $a=0,175$, la fel ca în primul caz dar, celălalt parametru b_2 se determină prin metoda celor mai mici pătrate (CMMP) monodimensională [7]:

$$b_2 = \arg(\min \{ \sum_{k=1}^{40} (p(k) - 0,175 k^{-b})^2 \}) \quad (9)$$

în care $p(k)$ reprezintă cele 40 de valori numerice ale probabilităților, determinate experimental și reprezentate în figura 2, iar modelul din (5) este:

$$p_2(k, b_2) = 0,175 k^{-b_2} \quad (10)$$

Erorile $e(k)$, de reprezentare a datelor experimentale $p(k)$ prin modelul $p_2(k)$ din (10), sunt exprimate prin diferențele $e(k) = p(k) - p_2(k, b_2)$. Aceste diferențe sunt funcții de parametrul b_2 care încă nu a fost calculat. Suma pătratelor erorilor $V(b_2)$, în acest caz, devine o funcție de b_2 :

$$V(b_2) = \sum_{k=1}^{40} (p(k) - p_2(k, b_2))^2 = \sum_{k=1}^{40} (p(k) - 0,175 k^{-b_2})^2 \quad (11)$$

Din graficul acestei funcții, prezentat în figura 3, se observă că minimum se obține pentru $b_2=4$.

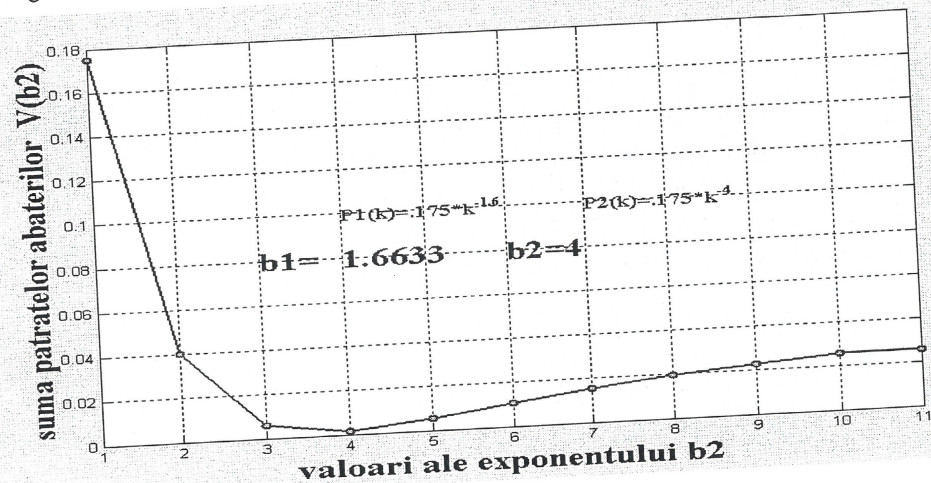


Figura 3. Graficul evoluției sumei abaterilor pătratice $V(b_2)$ în funcție de valorile parametrului

Pentru această valoare, suma pătratelor erorilor, în studiul de caz, rezultă:

$$SPE2 = \sum_{k=1}^{40} (p(k) - 0,175 k^{-4})^2 = 0,03 \quad (12)$$

Metoda 3 de estimare. În acest ultim caz, se determină prin metoda CMMP bidimensională, ambii parametri (atât a, cât și b):

$$[a_3, b_3] = \arg(\min \{ \sum_{k=1}^{40} (p(k) - a_3 k^{-b_3})^2 \}) \quad (13)$$

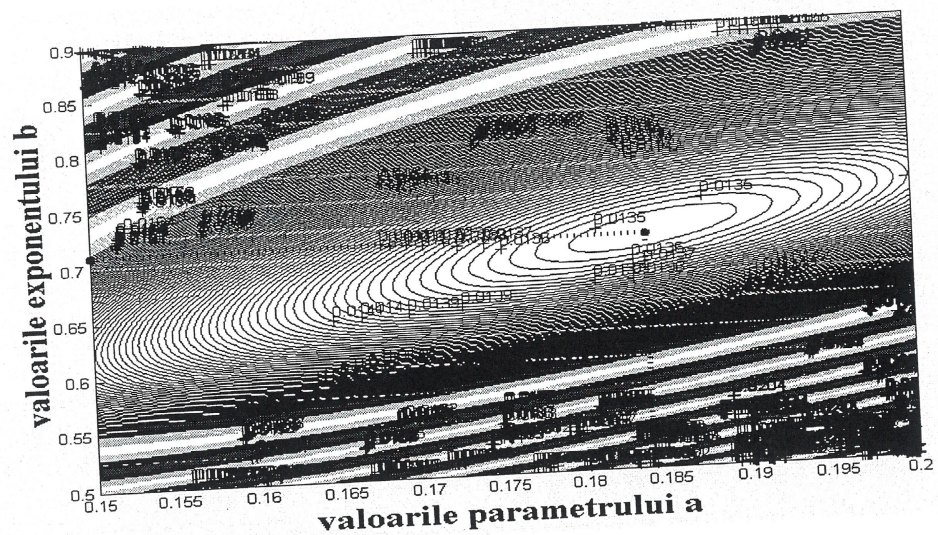


Figura 3. Graficul evoluției liniilor de izonivel ale sumei abaterilor pătratice $V(b_3, a_3)$ în funcție de valorile parametrilor a_3 și b_3

În acest caz, trebuie minimizată suma pătratelor, care este funcție de doi parametri:

$$V(a_3, b_3) = \sum_{k=1}^{40} (p(k) - a_3 k^{-b_3})^2 \quad (14)$$

$$a_3=0,71 \quad , \quad b_3=0,184$$
$$\text{SPE3} = \sum_{k=1}^{40} (p(k) - 0,71k^{-0,184})^2 = 0,004$$


3. Metode de analiză a similitudinii și importanței nodurilor rețelei

Cea mai simplă definiție a similitudinii este un echivalent al distanței Hamming și anume distanța exprimată prin numărul n_{ij} de vecini similari (adică, comuni) ai nodurilor i și j . Dacă nu există nici un vecin comun distanța dintre noduri se poate considera maximă. În caz contrar, dacă toți vecinii celor două noduri sunt comuni (adică și nodul i și nodul j au aceiași vecini) distanța dintre noduri se poate considera nulă. Aceasta reprezintă a doua cale de evaluare a similitudinii.

A treia cale constă în analiza statistică a vectorilor caracteristici prin determinarea probabilității de apariție în rețea a nodurilor de diverse grade. Din acest punct de vedere, există o regulă empirică, valabilă pentru rețelele complexe, în general, dar cu precădere este valabilă pentru rețelele complexe din domeniul financiar bancar: nodurile de grad mic au o probabilitate mare de apariție (și viceversa) .3.1 Evaluarea distanței între noduri vecine prin abaterea medie pătratică

Revista Română de Informatică și Automatică, vol. 15, nr. 4, 2005

propune ca măsura a similitudinii, abaterea medie pătratică X_{ij} între elementele celor doi vectori caracteristici, A_i și A_j de dimensiune N , aferenți nodurilor comparate:

$$X_{ij} = \frac{\sum_{k=1}^N (a_{ik} - a_{jk})^2}{N_v} \quad (15)$$

în care N_v reprezintă suma gradelor celor 2 noduri (adică suma totală de vecini N_i și N_j ai celor 2 noduri ale rețelei), iar a_{ik} și a_{jk} sunt elemente ale matricei de adiacență de pe liniile i și j ale acesteia. Deci, a_{ik} și a_{jk} sunt elemente ale celor doi vectori caracteristici, aferenți nodurilor comparate. Din relația (5), se observă că, în cazul în care cei doi vectori caracteristici, A_i și A_j , sunt egali (adică vecinii sunt toți comuni și $a_{ik}=a_{jk}$ pentru $k=1, 2, \dots, N_v$) gradul de asemănare $X_{ij}^*=0$. Dacă nodurile i și j nu au nici un vecin comun, atunci distanța este maximă, $X_{ij}^*=1$, deoarece suma pătratelor

$$Sp = \sum_{k=1}^N (a_{ik} - a_{jk})^2 \quad (16)$$

este egală cu N_v .

Forma vectorială a expresiei de calcul a sumei pătratelor este:

$$Sp = (A_i - A_j) \times (A_i - A_j)' \quad (17)$$

Pe baza acestei relații, se poate calcula distanța:

$$X_{ij} = \frac{S_p}{N_v} \quad (18)$$

În figura 6, sunt prezentate rezultate ale calculului abaterii medii pătratice între vectorii caracteristici aferenți rețelei a cărui graf este reprezentat în aceeași figură.

3.2 Metoda evaluării distanței dintre mulțimile de vecini ai nodurilor comparate

Distanța dintre două mulțimi (seturi) și S_j de vecini ai nodurilor i respectiv j se poate exprima prin raportul dintre numărul vecinilor necomuni și numărul total de vecini (suma vecinilor primului nod și a vecinilor nodului secund) cu relația:

$$X_s = \frac{(\text{numarul de vecini necomuni})}{(\text{totalul de vecini})} \quad (19)$$

Se observă din relația (19) că, în cazul în care cele două mulțimi coincid (toți vecinii sunt comuni), distanța este nulă $X_s=0$. Dacă nodurile i și j nu au nici un vecin comun, atunci distanța $X_s=1$. În figura 5, este prezentat un caz în care primul nod are 8 vecini, iar celalalt - 11 vecini (totalul fiind 16).

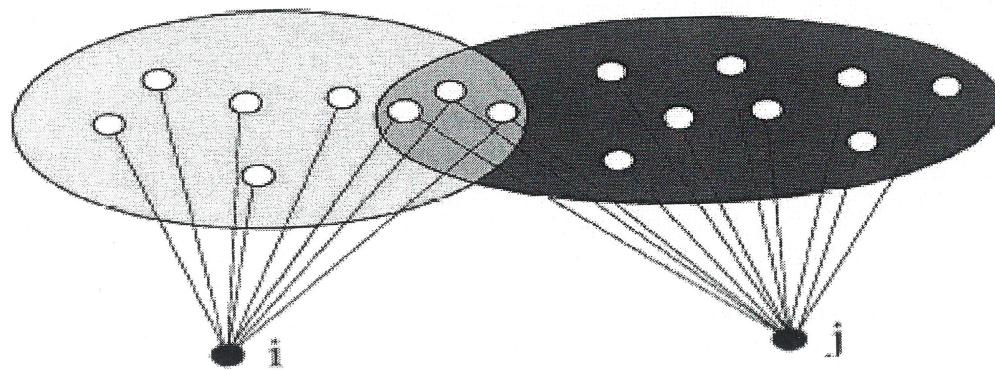


Figura 5. Mulțime formată din 16 vecini ai nodurilor i (care are în total 8 vecini) și j (care are în total 11 vecini), din care 3 vecini sunt comuni, iar 13 necomuni

Reuniunea celor două mulțimi de vecini este compusă din 16 noduri, iar intersecția conține 3 vecini comuni. Aplicând relația (19) rezulta distanța:

$$X_s = \frac{(16-3)}{(16+3)} = \frac{13}{19} \approx 0,68$$

3.3. Metoda evaluării distanței dintre vectorii caracteristici ai nodurilor comparate

În continuare, transpunem ideile de mai sus într-o reprezentare vectorială. Pentru subgraful din figura 3, se pot scrie vectorii caracteristici aferenți nodurilor i și j ale subgrafului care corespund liniilor A_i și A_j ale matricei A de adiacență a rețelei din care provine acest subgraf. Aceștia sunt doi vectori a căror dimensiune este egală cu numărul de noduri ale grafului rețelei și sunt descriși de următoarele expresii:

$$k=1;k=2;k=3;k=4;k=5;k=6;k=7;k=8;k=9;k=10;k=11;k=12;k=13;k=14;k=15;k=1$$

$$A_j = [0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]$$

$$A_i = [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$$

Cunoscutul metalimbaj de programare Matlab, conține metainstrucțiuni pentru calcule matriciale element cu element ale matricelor, fapt pentru care vom prezenta, în continuare, modul matricial de calcul a distanței X_s exprimată de relația (19). Vom atenționa cu semnul ! faptul că operațiile matriciale care urmează după acest semn, se efectuează element cu element. Produsul termen cu termen al celor doi vectori A_i și A_j de mai sus, este egal cu vectorul A_{ij} al vecinilor comuni:

$$A_{ij} = (A_i \times A_j) = [0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \quad (20)$$

În mod similar, se pot calcula vectorii vecinilor necomuni A_{i0} și A_{j0} prin operații de scădere a vectorului vecinilor comuni din vectorii caracteristici A_i și A_j corespunzători celor două linii din A :

$$A_{j0} = (A_j - A_{ij}) = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]$$

$$A_{i0} = (A_i - A_{ij}) = [1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0] \quad (21)$$

Folosind vectorii exprimați de relațiile (20) și (21), se obține următoarea relație de calcul matricial, element cu element, a distanței dintre vecinii celor două noduri:

$$X_{ij} = \frac{(A_{i0} + A_{j0}) \times (A_{i0} + A_{j0})'}{(A_{i0} + A_{j0}) \times (A_{i0} + A_{j0})' + 2 \cdot (A_{ij} \times A_{ij})'} \quad (22)$$

în care s-a notat cu $(A_{i0} + A_{j0})'$ vectorul $(A_{i0} + A_{j0})$ transpus.

Calculul distanței între vecinii nodurilor	9
Rezultatul calculului distanței între A_i, A_j	8
$X_{ij} = 0.68421$	
Gradul $d_i=8$ pt. nodul i și $d_j=11$ pt. nodul j	7
$X_{ij} = \text{numerator/numitor}$	6
$\text{numitor} = \text{numerator} + 2 \cdot A_{ij} \cdot A_{ij}'$	5
$\text{numerator} = (A_{i0} + A_{j0}) \cdot (A_{i0} + A_{j0})'$	4
$A_{ij} = A_i \cdot A_j; A_{j0} = A_j - A_{ij}; A_{i0} = A_i - A_{ij}$	3
$A_i = [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0]$	2
$A_j = [0 \ 0 \ 0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]$	1

Figura 6. Programul Matlab și rezultatul pentru calculul matricial cu relația (12) a distanței X_{ij} dintre vectorii caracteristici ai nodurilor i și j din rețeaua prezentată în figura 3

În cazul utilizării relației (19), este necesară o investigație prealabilă a matricei de adiacență pentru identificarea mulțimilor folosite în relația respectivă. Deoarece sursa inițială de informații este aceeași, cele două relații (19) și (22) dau același rezultat. În figura 6, este prezentat atât programul Matlab, cât și rezultatul obținut cu relația (22). Se observă că rezultatul obținut este $X_{ij}=0,68421$ și, deci, este același cu rezultatul obținut prin calcul cu relația (19): $X_s=0,68$.

3.4. Metoda evaluării coeficientului de corelație între vectorii caracteristici ai nodurilor rețelei [2]

Dacă rețeaua are N noduri j , atunci dimensiunea matricei de adiacență A este xN . În acest caz, distanța dintre noduri poate fi evaluată pe baza mediei și varianței valorilor de-a lungul liniilor matricei de adiacență (adică a vectorilor caracteristici). Pe baza acestora se calculează coeficienții de corelație R_{ij} dintre vectorii caracteristici i și j care poate lua valori între -1 și $+1$. Dacă cei doi vectori caracteristici nu au nici un vecin comun $R_{ij}=-1$, iar în cazul extrem opus - când toți vecinii ambelor noduri sunt comuni $R_{ij}=1$. Această metodă implică parcurgerea câtorva etape preliminare de prelucrare a vectorilor caracteristici. Pentru a deosebi vectorii de scalari, se folosesc litere îngroșate pentru identificatorii vectorilor. Metoda presupune o etapă preliminară de centrare, prin valori medii și, respectiv, de normare, prin dispersii, a vectorilor comparați.

Centrarea și normarea vectorilor caracteristici. Pentru vectorul A_i media m_i este scalarul egal cu suma elementelor (calculată ca produs al vectorului cu transpusa acestuia $A_i \times A_i'$) împărțită cu N :

$$m_i = \frac{A_i \cdot A_i'}{N} \quad (23)$$

Valorile centrate A_{ci} ale vectorului A_i aferent nodului i se obțin cu relația:

$$A_{ci} = A_i - m_i \quad (24)$$

Pătratul dispersiei vectorului este scalarul egal cu suma pătratelor elementelor vectorului centrat, împărțită cu N :

$$\sigma_i^2 = \frac{A_{ci} \cdot A_{ci}'}{N} \quad (25)$$

Normarea valorilor centrate, conținute de vectorul A_{ci} se face prin împărțirea lor la varianta σ_i , rezultând vectorul caracteristic normat A_{0i} al nodului i :

$$A_{0i} = \frac{A_{ci}}{\sigma_i} \quad (26)$$

În mod similar cu relația (26), se poate determina vectorul caracteristic cu valori centrate și normate pentru nodul j :

$$A_{0j} = \frac{A_{cj}}{\sigma_j} \quad (27)$$

Calculul coeficientului de corelație. Coeficientul de corelație R_{ij} între vectorii i și j este un scalar și se calculează împărțind prin N produsul dintre vectorul caracteristic (26) și vectorul (27) transpus:

$$R_{ij} = \frac{A_{0i} \cdot A_{0j}'}{N} = \frac{A_{ci} \cdot A_{cj}'}{\sigma_i \cdot \sigma_j \cdot N} \quad (28)$$

Coefficientii de corelație R între vectorii caracteristici ai rețelei

$$R(A) = \begin{bmatrix} 1 & -0.091287 & 0.3 & 0.091287 & -0.4 & -0.4 & 0.6455 \\ -0.091287 & 1 & -0.091287 & -1 & -0.091287 & -0.091287 & -0.4714 \\ 0.3 & -0.091287 & 1 & 0.091287 & -0.4 & -0.4 & 0.6455 \\ 0.091287 & -1 & 0.091287 & 1 & 0.091287 & 0.091287 & 0.4714 \\ -0.4 & -0.091287 & -0.4 & 0.091287 & 1 & 0.3 & -0.2582 \\ -0.4 & -0.091287 & -0.4 & 0.091287 & 0.3 & 1 & -0.2582 \\ 0.6455 & -0.4714 & 0.6455 & 0.4714 & -0.2582 & -0.2582 & 1 \end{bmatrix}$$

Matricea de adiacență a grafului dat este A

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

GRAFUL UNEI MICI REȚELE DE TIP COMUNITAR

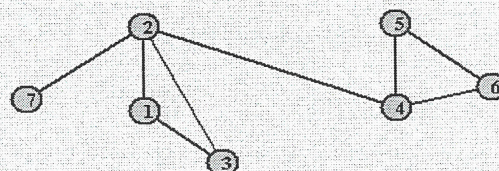


Figura 7. O rețea mică de legături între 7 agenți economici (segment de rețea complexă)

Se poate observa că atunci când toți vecinii sunt comuni, atunci ($A_{ci} = A_{cj}$), iar ($\sigma_i = \sigma_j = \sigma$). În acest caz, relația (28) devine:

$$R_{ij} = A_{0i} \times A_{0j} = \frac{A_{ci} \times A_{ci}'}{\sigma_i \times \sigma_j \times N} = \frac{A_{ci} \times A_{ci}'}{\sigma^2 \times N} \quad (29)$$

Înlocuind în (29) pe σ_i cu expresia sa din relația (25) se obține, în acest caz $R_{ij} = 1$ deoarece în (29) avem, ($A_{ci} \times A_{ci}' = (\sigma^2 \times N)$).

În cazul extrem opus, când nodurile nu au nici un vecin comun, $R_{ij} = -1$ deoarece ($A_{ci} \times A_{cj}' = -(\sigma^2 \times N)$). Cazul în care $R_{ij} = 0$ corespunde unui caz intermediar în care fie este nul unul din cei doi vectori A_{ci} sau A_{cj} , fie ambii sunt nuli. Această anulare o poate provoca operația de centrare prin valoarea medie m_i sau m_j a celor 2 vectori caracteristici comparați A_i și A_j .

În Matlab, există comanda cu mnemonicul Corrcoeff, care calculează toți coeficienții de corelație între coloanele unei matrice. Sintaxa comenzii este:

$$R_A = \text{Corrcoeff}(A) \quad (30)$$

și are ca argument o matrice oarecare A, între coloanele căreia se calculează coeficienții de corelație. Se returnează o matrice care pe poziția $R_{ij}(A)$ conține valoarea coeficientului de corelație dintre coloanele i și j ale matricei A. Matricea $R(A)$ care rezultă este simetrică, deoarece $R_{ij}(A) = R_{ji}(A)$. În figura 7, se prezintă o minirețea cu șapte noduri, matricea A de incidență a grafului rețelei și matricea corelațiilor între toți vectorii caracteristici ai rețelei calculată în Matlab, folosind instrucțiunea (20). Se observă că, întrucât vectorii A_2 și A_4 nu au vecini comuni, se obține: $R_{24} = R_{42} = -1$, iar corelațiile vectorilor caracteristici cu ei înșiși sunt egale cu unitatea și sunt reprezentate pe diagonala matricei $R(A)$ din figura 7.

3.5. Centralitatea de vector propriu [1]

O mult mai sofisticată versiune a aceleiași idei este așa numita centralitate de vector propriu, „eigenvector centrality”. În timp ce simplul număr de arce care se întâlnesc în nod exprimă centralitatea de grad a nodului, centralitatea de vector propriu consideră că nu toate arcele care se întâlnesc în nod sunt la fel de importante. Deci, se propune ca ponderea fiecărui vecin în evaluarea centralității nodului să fie diferită și în conformitate cu importanța lor. În general, conexiunile cu persoanele mai influente sunt mai valoroase.

Dacă notăm cu x_i centralitatea transformată a nodului i, se poate considera că efectul pe care îl are x_i este proporțional cu centralitățile cumulate ale tuturor celor n vecini ai nodului i:

$$x_i = \frac{1}{\lambda} \sum_{j=1}^n A_{ij} x_j \quad (31)$$

în care λ_i este o constantă. O asemenea formulare a centralității exprimă nivelul importanței nodului i în funcție de nivelul importanței tuturor vecinilor săi conform proverbului străvechi: „spune-mi cu cine te întâlnești ca să-ți spun cine ești”. În limbaj algebric, această ecuație exprimă o transformare liniară prin intermediul vectorului caracteristic, al centralităților x_i a tuturor vecinilor în vectorul x^* al centralităților vecinilor cumulate în nodul i :

$$x^* = \sum_{j=1}^n A_{ij} x_j \quad (32)$$

Dacă $x^* = \sigma x_i$, atunci λ_i se numește valoare proprie a centralității x_i și ecuația transformării liniare devine:

$$\lambda x_i = \sum_{j=1}^n A_{ij} x_j \quad (33)$$

Introducând un vector cu centralitățile tuturor nodurilor rețele $x = (x_1; x_2; \dots)$, se poate rescrie această relație într-o formă matricial-vectorială. Această ecuație matricială de transformare liniară prin matricea de adiacență a centralităților:

$$x^* = \lambda x = A x$$

în care x se numește vector propriu al matricei de adiacență A , iar λ se numește valoare proprie a vectorului x .

Centralitatea de vector propriu, exprimată prin vectorul propriu $x > 0$ al matricei de adiacență A și prin intermediul valorii proprii λ ține cont atât de numărul de vecini ai nodului, dar și de calitatea (sau de nivelul de importanță) acestor vecini. Dacă un nod posedă un mare număr de contacte cu vecini mediocri, această carență poate fi compensată de existența câtorva contacte de foarte înaltă calitate [3].

Proprietatea de coliniaritate a centralității de vector propriu. Această proprietate a vectorului propriu constă în aceea că orice alt vector $x^* > 0$ proporțional cu x constituie exprimarea, la altă scară, a centralității de vector propriu al aceleiași matrice de adiacență A și de aceeași valoare proprie λ . Această proprietate decurge din proprietatea de liniaritate a transformării liniare. Conform acestei proprietăți de liniaritate, pentru vectorul proporțional, $x^* = b x$, se poate scrie:

$$A x^* = b A x = b (A x) = b (\lambda x) = \lambda (b x) = \lambda x^*$$

din care rezultă că și x^* este vector propriu al matricei de adiacență A cu aceeași valoare proprie λ .

Exemplu pentru procedeul de calcul a vectorului propriu al matricei transformatoare A . Fie matricea

$$\text{transformării liniare } A = \begin{bmatrix} a_{11} & \dots & a_{12} \\ a_{21} & \dots & a_{22} \end{bmatrix}$$

Presupunem că vectorul $v = [r; s] = [r, s]$ (aici semnul „;” între elementele vectorului semnifică faptul că vectorul respectiv este de tip coloană, iar virgula între elemente semnifică „vector linie” și apostroful semnifică operația transpunerii) este vector propriu al matricei A , iar λ este valoarea proprie a acestuia. Deci, relația transformării liniare este, $\lambda v = A v$

care devine:

$$\begin{aligned} a_{11} \cdot r + a_{12} \cdot s &= \lambda r & a_{11} - \lambda & & a_{12} \cdot s &= 0 \\ & \} \text{ sau} & & & & \\ a_{21} \cdot r + a_{22} \cdot s &= \lambda s & a_{21} r & + & (a_{22} - \lambda) \cdot s &= 0 \end{aligned} \quad (34)$$

Deoarece vectorul $v = [r; s]$ este, prin definiție, diferit de zero, pentru sistemul (34) se poate scrie ecuația caracteristică a sistemului (35):

$$\text{Det} \begin{bmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{bmatrix} = (a_{11} - \lambda)(a_{22} - \lambda) - a_{21} a_{12} = 0 \quad (35)$$

în care, $\text{Det}()$ semnifică „determinantul matricei din paranteză”.

Valoarea proprie λ se calculează rezolvând ecuația caracteristică (36), iar această valoare se introduce în (35) și rezolvând-o se obține vectorul caracteristic al matricei A .

Aplicatie. Fie $A = \begin{bmatrix} 1 & \dots & 1 \\ 0 & \dots & 1 \end{bmatrix}$

Ecuția (36) devine:

$$(1 - \lambda)^2 = 0.$$

Soluția ecuației este $\lambda^* = 1$, iar după înlocuirea în ecuația (35) rezultă rădăcinile sistemului: $s^* = 0$ și r^* poate lua orice valoare pe axa absciselor. Astfel, vectorii proprii ai matricei A , din această aplicație, sunt de forma $v^* = [r^*; 0]$ și sunt geometric plasați pe axa absciselor.

Ecuția caracteristică (36) a matricei A se poate scrie sub forma matriceală:

$$\text{Det}(A - \lambda I) = 0 \quad (36)$$

în care I este matricea unitate (de aceeași dimensiune ca și A) iar $\text{Det}()$ semnifică „determinantul matricei din paranteză”.

Polinomul caracteristic al matricei transformatoare A . Deoarece partea stângă a ecuației (36) și paranteza din (37) are forma unui polinom în λ , determinantul expresiei din paranteza a ecuației (37),

$$\text{Det}(A - \lambda I), \quad (37)$$

se numește polinom caracteristic al matricei transformatoare A . În cazul rețelelor complexe, cu arce neorientate, matricea de transformare liniară este *matricea de adiacență* a rețelei.

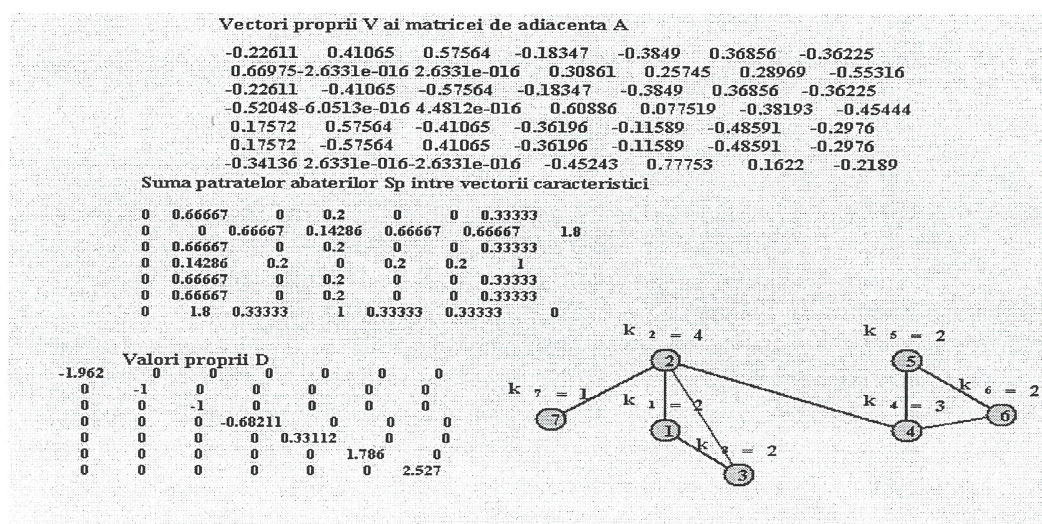


Figura 8. Rezultatele calculului vectorilor proprii ai matricei de adiacență V , abaterii pătratică între vectorii caracteristici Sp și gradele $k_1, \dots, k_7$ pentru nodurile grafului din figura 5

În figura 8, sunt prezentate trei tipuri de indicatori de centralitate, calculați pentru același graf: centralitate de vector propriu, centralitate de grad și indicator de abatere medie pătratică.

În încheiere, subliniem că centralitatea de grad se referă la nodul propriu-zis și este funcție numai de vecinii săi cei mai apropiați, iar centralitatea de vector propriu se exprimă în funcție de centralitățile tuturor celorlalte noduri din rețea, după cum rezultă din relația (32). Aceasta face mult mai dificilă interpretarea rezultatelor, în cazul unei rețele reale. Acest aspect este vizibil din figura 8 în care se prezintă, chiar pe graf gradele $k_1, k_2, \dots, k_7$ ale nodurilor, iar în partea stângă a figurii sunt afișate: matricea vectorilor proprii și matricea valorilor proprii. Aceste două matrici, notate V și respectiv D , au fost calculate în Matlab folosind instrucțiunea:

$$[V, D] = \text{eig}(A) \quad (38)$$

Având în vedere dificultățile de interpretare și de calcul, în aplicații, se utilizează mai rar centralitatea de vector propriu.

4. Concluzii

Estimarea parametrilor distribuției exponențiale a fost testată pe un studiu de caz prin trei metode. Metoda cea mai precisă s-a dovedit a fi metoda CMMP.

Dacă prima secțiune a lucrării furnizează informații despre modul în care se poate evalua probabilitatea existenței unor clienți valoroși ai băncii, în a doua secțiune se determină indicatori ai rețelei care permit să se precizeze care sunt acești clienți valoroși ai băncii.

Matricea de incidență a grafului rețelei complexe oferă posibilitatea formalizării reprezentării indicatorilor de centralitate. Pentru comoditatea analizei individuale sau în perechi a nodurilor, s-a introdus conceptul de vector caracteristic al nodului. Pe această bază, au fost propuși algoritmi pentru determinarea indicatorilor de centralitate prin calcul matriceal. Centralitatea de vector propriu se exprimă în funcție de centralitățile tuturor celorlalte noduri din rețea, după cum rezultă din relația (32). Aceasta face mult mai dificilă interpretarea rezultatelor, în cazul unei rețele reale.

Bibliografie

1. **HOLME, P.:** Congestion and Centrality în Traffic Flow on Complex Networks. În: Adv. Complex Syst. 6 (2003), pp. 163-176.
2. **NEWMAN, M. E. J.:** The Structure and Function of Complex Networks. În: SIAM Review 45, 2003, pp. 167-256.
3. **RÎPEANU, M., I. FOSTER, A. IAMNITCHI:** Mapping the Gnutella network: Properties of Large-scale Peer-to-Peer Networks and Implications for System Design. În: IEEE Internet Comput. 6 (2002), pp. 50-57.
4. **BARABASI, A.L., R. ALBERT, H. JEONG, G. BIANCONI:** Power-law Distribution of the World Wide Web. În: Science 287 (2000), p. 2115.
5. **STĂNCIULESCU, FL.:** Modeling of High Complexity Systems With Applications, WIT-press, 2005.
6. **STĂNCIULESCU, FL. :** A Fuzzy Expert System for Simulation and Control of High Complexity Systems, The 5th Congress on Modeling and Simulation, Paris, 2004.
7. **TERTIȘCO, M., P. STOICA:** Identificarea sistemelor, Ed. Academiei Române, 1980.