MODEL TERMODINAMIC AL DIFUZIEI INFORMAȚIEI ȘTIINȚIFICE

AUREL AVRAMESCU
membru titular al Academiei Republicii Socialiste România

Comunicare prezentată în ședința Secției de științe tehnice, din 30 iunie 1977

THERMODYNAMIC PATTERN OF SCIENTIFIC INFORMATION DIFFUSION. Following the common sense meaning of diffusion, the author succeeded in discovering the numerable bibliographic parameters corresponding to temperature, space and flow in the heat conduction process in solids. These are : the interest of new authors for previous articles, measured by the number of references ; the space of ordered readers/virtual authors who receive, accumulate and transfer information. In this way the local information balance leads to formulation of flow and energy laws and consequently to Fourier's differential equation, whose solution for an impulsory source placed in the space origin happens to be the family of Gaussian bell curves, also describing dispersion probability of random measurement errors. The indicial solution for a permanent unit source is given by the complementary error function. In both cases the mean square space coordinate is tightly bound to time, a relation which allows to obtain Bradford's law of articles scattered in scientific journals. Presuming a given (temperature) level in works treating the same problem and an initial source with first growing and then slowly damped information output the author succeeded in reproducing the strange core behaviour and the empirically stated saturation trends of article number ("Groos droop").

Simple presumptions allow further to get the analytical expression of the citation distribution of earlier papers in new ones as a function of age, which exhibits, beyond the normal exponential citation obsolescence of older works, a peculiar "bounce" for recently appeared articles,outnumbering six times the exponential law. Instead of accepting a consequent bipartition of recent works in "archival" and "fresh" ones, the theory offers a single function describing the stated behaviour in the whole age range, making a criterionless bipartition unnecessary.

Another much discussed problem by most distinguished authors concerning a sound definition of information entropy can be cleared up by writing wellknown thermodynamic formulae of local energy and entropy density changes. As known, in "bond graphs" temperature multiplied by entropy flow gives the thermal power, explaining why entropy is responsible for process dynamics. It is clearly shown how on one hand sources contribute to decrease system entropy by negentropy output and on another, information transfer is bound to entropy rise in all space elements involved. So, information flow introduces order and organization in systems, braking their normal or accelerated decay. In this way by analysing local entropy flow we gain insight in information transfer processes, as for instance in case of the mentioned "immediacy effect", which show that the entropy drop is even larger then temperature fall in the neighbourhood of sources, and explain the much increased interest of new authors for very recently published papers.

The physical diffusion model proves to be efficient in describing many aspects of the information dissemination process.

Transferul și răspîndirea cunoștințelor științifice se pot asimila cu un proces de diseminare biologică, în care sădirea și incubarea sînt urmate de creștere și fructificare, în cicluri succesive. Mai puțin complex ar fi un model fizic, tratabil matematic, similar cu difuzia informației științifice, care ar putea confirma caracterul presupus logic al răspîndirii științei. În fapt, din ridicări statistice în domeniul bibliometriei, au rezultat o serie de legități surprinzătoare privind ritmul constant, dar foarte rapid, al creșterii publicațiilor științifice în ultimele trei secole[16], răspîndirea unei cercetări eminente în progresie geometrică în revistele de specialitate (legea Bradford) [8] sau distribuția productivității autorilor într-un grup dat (legea Lotka [14]), al căror număr scade invers proporțional cu o putere a numărului de lucrări elaborate într-un timp dat. Autorul și-a pus, încă acum 15 ani, întrebarea dacă procesul de creștere și de difuzie a informațiilor științifice nu s-ar putea asimila cu fenomenele de difuzie din fizică, dominate de teoria fundamentată de J. B. Fourier acum 155 ani și amplu dezvoltată pînă în zilele noastre de către savanți de mare renume. Asimilarea a reușit în anumite condiții, dealtfel puțin restrictive, prezumțiile și rezultatele fiind publicate într-o serie de lucrări, favorabil comentate de către specialiști în domeniu. Rezultatul esențial al acestei modelări se concretizează prin descrierea și explicarea precisă și plauzibilă a legităților bibliometrice stabilite empiric pe baza a numeroase statistici [3] — [6]. Comentariile au incitat autorul să extindă teoria la aspectele termodinamice mai generale, rezultînd, în cadrul prezumțiilor făcute, *o nouă definiție a entropiei și energiei informaționale*, ceea ce ar putea să redeschidă controversa privind conținutul acestor mărimi în teoria matematică a informației și în termodinamică [9], [15].

În cele ce urmează, se prezintă în mod succint rezultatele principalelor faze de fundamentare a teoriei noi, plecînd de la prezentarea datelor bibliometrice ridicate statistic la stabilirea parametrilor numerabili ai procesului difuziv și la stabilirea relațiilor dintre aceștia, trecînd apoi la asimilarea cu teoria analitică a difuziei căldurii în corpuri solide și terminînd cu interpretarea corespunzătoare a rezultatelor deduse teoretic, inclusiv a consecințelor privind dilema entropiei informaționale.

DATE ȘI RELAȚII BIBLIOMETRICE

Ca în oricare cercetare a fenomenelor naturale, din multitudinea de aspecte care pot caracteriza progresul științei, se aleg strict numai parametrii de stare numerabili care permit observarea obiectivă și o tratare statistică. Din fericire, prin caracterul de masă al majorității parametrilor observabili privind transferul de informații științifice, în special numărul de autori, lucrări, reviste, referințe și citări, legile statisticii matematice și teoria probabilităților se pot aplica în condiții similare cu cele din teoria analitică a termodinamicii.

Numărul revistelor științifice j în care se publică rezultate notabile de cercetare crește, de trei secole, cu o rată constantă de dublare în fiecare 15 ani, recuperînd reducerile de ritm după mari războaie și alte evenimente frenatoare [16]. Acest ritm deosebit de rapid se poate media cu o precizie de $\pm 1{,}5\%$ prin expresia exponențială :

$$j(t) = j_0 \exp (t/T_j), \tag{1}$$

unde j_0 înseamnă numărul de reviste la începutul numărătorii, t timpul în ani, iar constanta de timp $T_j = 22$ ani. De notat că între timpul de dublare, constanta de timp T și creșterea anuală procentuală p, există relațiile aproximative : $T = t_2/0{,}7 = 100/p$.

Numărul articolelor în revistele științifice crește cu o rată constantă și mai mare, anume cu 7% anual, dublîndu-se în numai 10 ani :

$$x(t) = x_0 \exp (t/T_x), \tag{2}$$

notațiile corespunzînd cu cele din relația (1). Raportul $T_j/T_x = 1{,}5$ are o deosebită stabilitate, exprimînd creșterea cu ritm constant a articolelor din reviste în decursul anilor.

După ultimele statistici efectuate cu calculatorul electronic, lucrările științifice actuale au în medie 16 ± 6 referințe bibliografice, cele care au mai puțin de 10 considerîndu-se slab documentate, iar cele cu peste 22 avînd caracter de sinteză. Prin referiri, autorii evidențiază utilizarea unor informații „relevante" din literatura de specialitate, pentru conceperea lucrărilor lor de cercetare, care conțin, bineînțeles, și contribuții originale.

Interesantă apare *distribuția citărilor multiple* ale unei lucrări, găsindu-se că 5 sau 10 citări, în lucrările anterioare reprezintă 1, respectiv 0,1% din cele citate o singură dată. Două statistici privind frecvența lucrărilor multiplu citate [17] concordă în confirmarea unei distribuții armonice cubice a lucrărilor x cu numărul m al citatelor pe lucrare : $x(m) = m^{-3}$. Or, suma infinită a seriei respective este dată de funcția Zeta (Riemann), $\zeta(3) = 1{,}2020569\ldots$, iar suma primilor 10, respectiv 20 termeni, reprezintă o scădere de numai 0,37, respectiv 0,1% din această valoare. Astfel ar rezulta că numărul citatelor multiple ar majora cu numai 20% numărul lucrărilor o dată citate, permițînd deducerea numărului de citate din numărul total de lucrări.

Autorul a mai arătat că nu numai numărul, dar și *repartiția în timp a citărilor* primite de o lucrare contribuie esențial la stabilirea valorii ei [1]. Exceptînd lucrările geniale, a căror citare continuă să crească exponențial în timp, pentru majoritatea lucrărilor multiplu citate se înregistrează o distribuție mai rapid ascendentă și mai încet descendentă a numărului de citate, similară cu unda normală de șoc a supratensiunilor atmosferice în instalațiile electrice de înaltă tensiune. O valoare mai mare trebuie atribuită lucrărilor la care distribuția citatelor se amortizează mai încet.

Dar, pentru propășirea științei, soarta unor lucrări individuale nu poate avea o pondere notabilă în sine, ci numai efectul colectiv al masei lucrărilor științifice. În acest sens, repartiția citărilor după vechimea articolelor și cea a referirilor din apariții recente la lucrări anterioare, pot să caracterizeze în mod mult mai cuprinzător forța propulsantă a științei. Această recunoaștere constituie momentul fundamental al prezentei cercetări, deoarece îndreaptă atenția asupra datelor statistice respective.

Legăturile dintre lucrări anterioare și cele apărute în cursul ultimului an se relevează în mod precis. La sfîrșitul anului există cu 7% mai multe articole ca la începutul lui. În 40% din acestea lucrările anterioare nu se citează de loc, 50% — o singură dată, 12% — de două ori, 6% — de trei ori ș.a.m.d. Pentru cele 60% de articole-surse citate, în figura 1 se arată legăturile cu cele prezente, manifestate prin referințe și citate. Prin natura lucrurilor, numărul de citate este riguros egal cu numărul referinţelor. Este interesant să se observe concentrarea acestor legături pe primele două lucrări-surse, care au primit 6, respectiv 4 citări. Acestea au stîrnit deci un interes deosebit faţă de celelalte, manifestat prin referiri multiple. Ele se pot considera aflate la un potenţial mai ridicat de la care porneşte un flux mai bogat de informaţii înspre autorii/lucrările mai recente. Ordonînd lucrările-sursă la stînga, în ordinea descrescîndă a citărilor şi trecînd de la o distribuţie discretă la una continuă (ceea ce se justifică pentru un număr foarte mare de lucrări), rezultă curba de potenţial avînd forma binecunoscută de pîlnie [3].

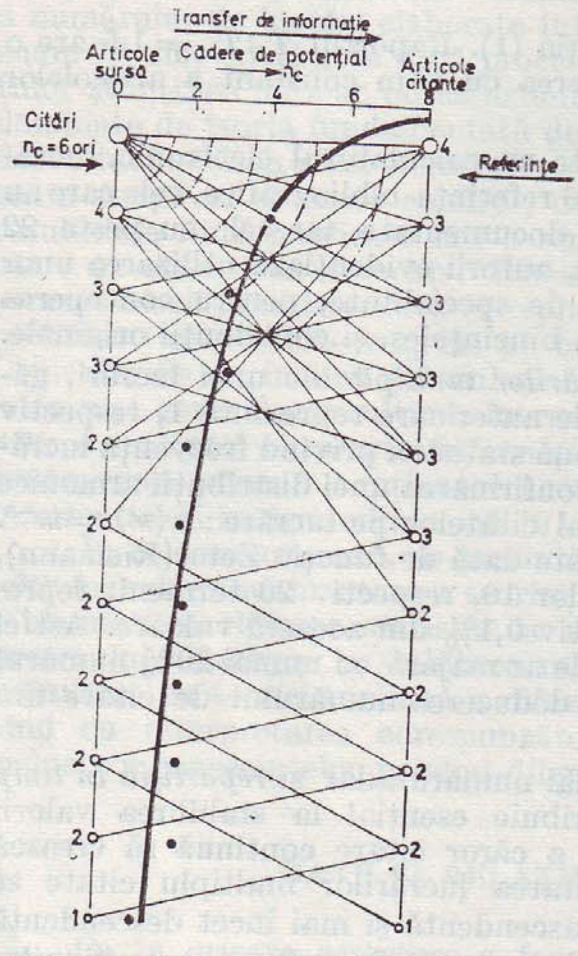

Fig. 1. — Cîmpul potenţial al lucrărilor multiplu citate, care transferă informaţii pentru elaborarea unor lucrări noi, în care primele apar ca referinţe bibliografice.

Distribuţia în vechime a lucrărilor citate a fost stabilită de către calculatorul Institutului de Informare Ştiinţifică (ISI) din Philadelphia S.U.A., care editează evidenţa trimestrială ,,Science Citation Index'' [11]. Tabloul ce rezultă din figura 2 este surprinzător. Pe cînd evaluări anterioare au indicat o scădere exponenţială simplă cu vechimea articolelor citate (dreapta a), calculatorul a pus în evidenţă un interes foarte mare în plus pentru lucrări foarte recente (curba d) care dă la vîrf valori de 6 ori mai mari.

APLICAREA TEORIEI DIFUZIEI

Materialul faptic înfăţişat, dar mai ales curba curioasă a obsolescenţei lucrărilor, pune problema dificilă de a explica aceste constatări surprinzătoare pe baza unei teorii consistente. Recurgerea la teoria analitică a difuziei se impune în mod natural, ca şi pentru răspîndirea epidemiilor. O primă încercare de asimilare a autorului [2] a dus la deducerea din teorie a legii lui Bradford [8], care stabileşte o progresie geometrică pentru răspîndirea în reviste a unei bibliografii privind o descoperire dată.

Condiţiile de asimilare au fost următoarele :

a) Informaţia conţinută într-o lucrare publicată şi difuzată ajunge la un număr mare de cititori interesaţi, care comunică şi între ei pe diferite căi, astfel că masa lor se poate asimila cu o mulţime de elemente discrete în contact conductiv.

b) Între surse şi receptori se stabileşte un flux de informaţii propulsat de către gradientul de potenţial al interesului care îl difuzează în spa-

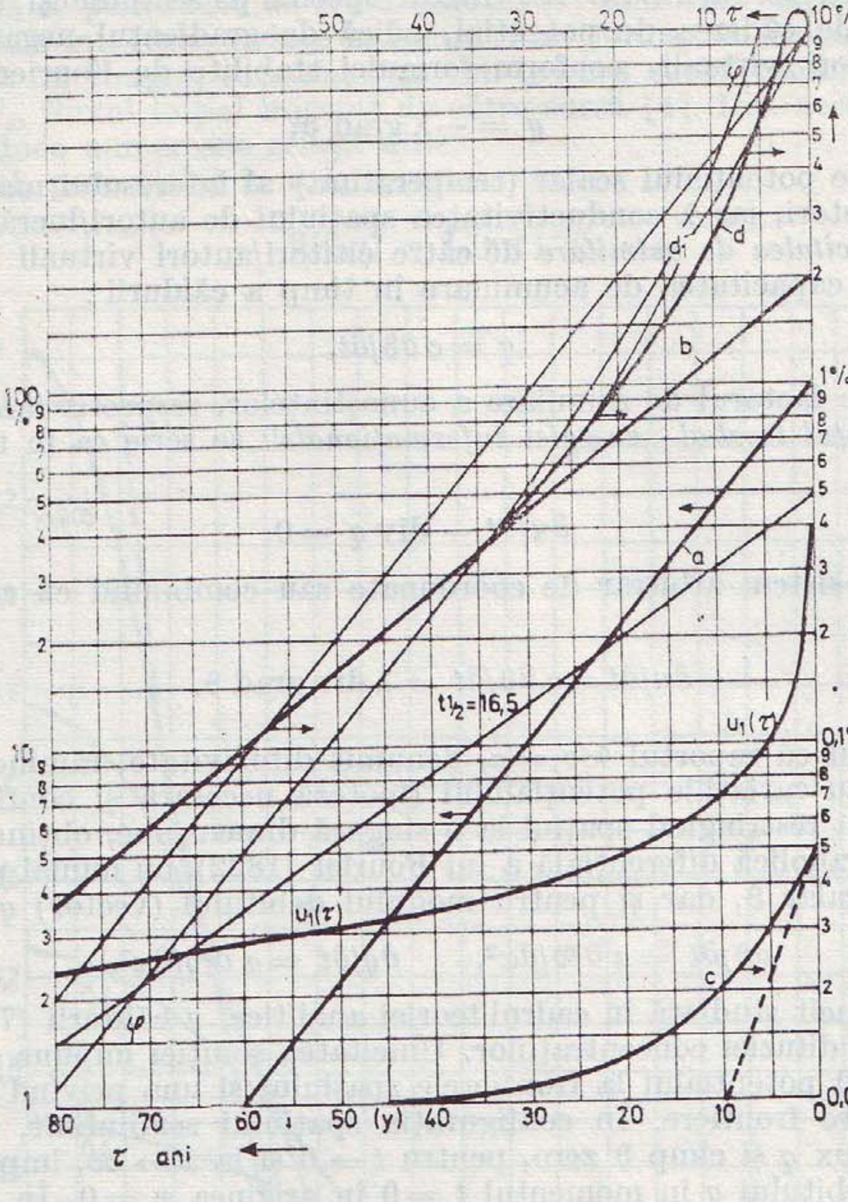

Fig. 2. — Obsolescența lucrărilor științifice citate în funcție de vechimea apariției (τ ani), rezultată din statistici ISI :

a — rata lucrărilor citate într-un an, mai vechi decît τ ;
b — rata lucrărilor mai vechi decît 40 ani, citate într-un an ;
φ — pontențialul dedus al interesului pentru lucrări anterioare ;
c — rata lucrărilor mai recente decît 40 ani ;
d — rata cumulată a citărilor pe an ;
d_1 — aproximația inițial liniară a curbei *d*.

țiul compact, omogen și izotrop al receptorilor : cititori, respectiv autori virtuali.

c) Informația primită de receptori, parțial se acumulează, ridicînd nivelul științific, parțial se retransmite altora, dar nu se pierde, ceea ce corespunde legii de conservare a energiei, din fizică.

În aceste condiții, formulările devin identice cu cele corespunzătoare difuziei fizice.

Debitul de informații sau fluxul specific pe secțiune și timp va fi vehiculat de căderea de potențial, adică de gradientul negativ, înspre cititori/autori virtuali, conform formulei stabilite de Fourier :

$$q = -\lambda \operatorname{grad} \vartheta, \tag{3}$$

unde ϑ este potențialul scalar (temperatura) al interesului manifestat de către receptori, iar λ conductivitatea spațiului de autori/lucrări [4].

Capacitatea de asimilare de către cititori/autori virtuali se exprimă similar cu capacitatea de acumulare în timp a căldurii :

$$q = c\, \partial\vartheta/\partial t, \tag{4}$$

unde c este factorul de asimilare a cunoștințelor, respectiv informațiilor.

Bilanțul local al „energiei informaționale" se scrie ca în termodinamică :

$$\partial u/\partial t - \operatorname{div} q = 0 \tag{5}$$

pentru un sistem arbitrar de coordonate sau combinînd cu relațiile (3) și (4)

$$\partial u/\partial t = c\, \partial\vartheta/\partial t = \lambda \operatorname{div} \operatorname{grad} \vartheta. \tag{6}$$

Presupunînd că raportul $\lambda/c = a$, denumit difuzivitate, rămîne constant în raport cu variațiile potențialului (ipoteză necesară și confirmată de rezultate) și restrîngînd spațiul la o singură dimensiune, obținem celebra ecuație parabolică diferențială a lui Fourier (1822), nu numai pentru potențialul scalar ϑ, dar și pentru modulul debitului (vector) q :

$$\partial\vartheta/\partial t = a\, \partial^2\vartheta/\partial x^2, \qquad \partial q/\partial t = a\, \partial^2 q/\partial x^2, \tag{7}$$

ecuație adîncit studiată în cadrul teoriei analitice a căldurii [7], precum și în teoria difuziei concentrațiilor. Unicitatea soluției impune două condiții privind potențialul la frontierele spațiului și una privind repartiția inițială între frontiere. În configurația spațiului semiinfinit, fie aceste condiții : flux q și cîmp ϑ zero, pentru $t \to 0$ și la $x \to \infty$, impuls unitar Dirac al debitului q în momentul $t = 0$ în originea $x = 0$. În acest caz, soluția pentru debitul q (în automatică funcția-pondere a sistemului) se poate deduce ușor cu ajutorul transformației Laplace [10], fiind :

$$q(x,t) = \frac{cI_0\, x}{2(\pi\, at)^{1/2}} \exp\left(-\frac{x^2}{4at}\right). \tag{8}$$

Soluția corespunzătoare pentru cîmpul potențial este funcția erorilor aleatorii a lui Gauss :

$$\vartheta(x, t) = \frac{I_0}{2(\pi\, at)^{1/2}} \exp\left(-\frac{x^2}{4at}\right), \tag{9}$$

sau, introducînd timpul numeric $y = 4\, at$,

$$\vartheta(x,y) = I_0(\pi y)^{-1/2} \exp\left(-x^2/y\right). \tag{9a}$$

Proba se poate face derivînd ecuația (9) conform relației (3).

Curbele integrale, redate în figura 3 au forma de clopote care se aplatizează cu timpul, suprafața de sub oricare curbă rămînînd riguros egală cu I_0, fluxul inițial injectat de către sursă [4]. Din această soluție se pot deduce numeroase relații utile.

Astfel, potențialul în origine este :

$$\vartheta(0,y) = I_0(\pi y)^{-1/2}, \tag{9b}$$

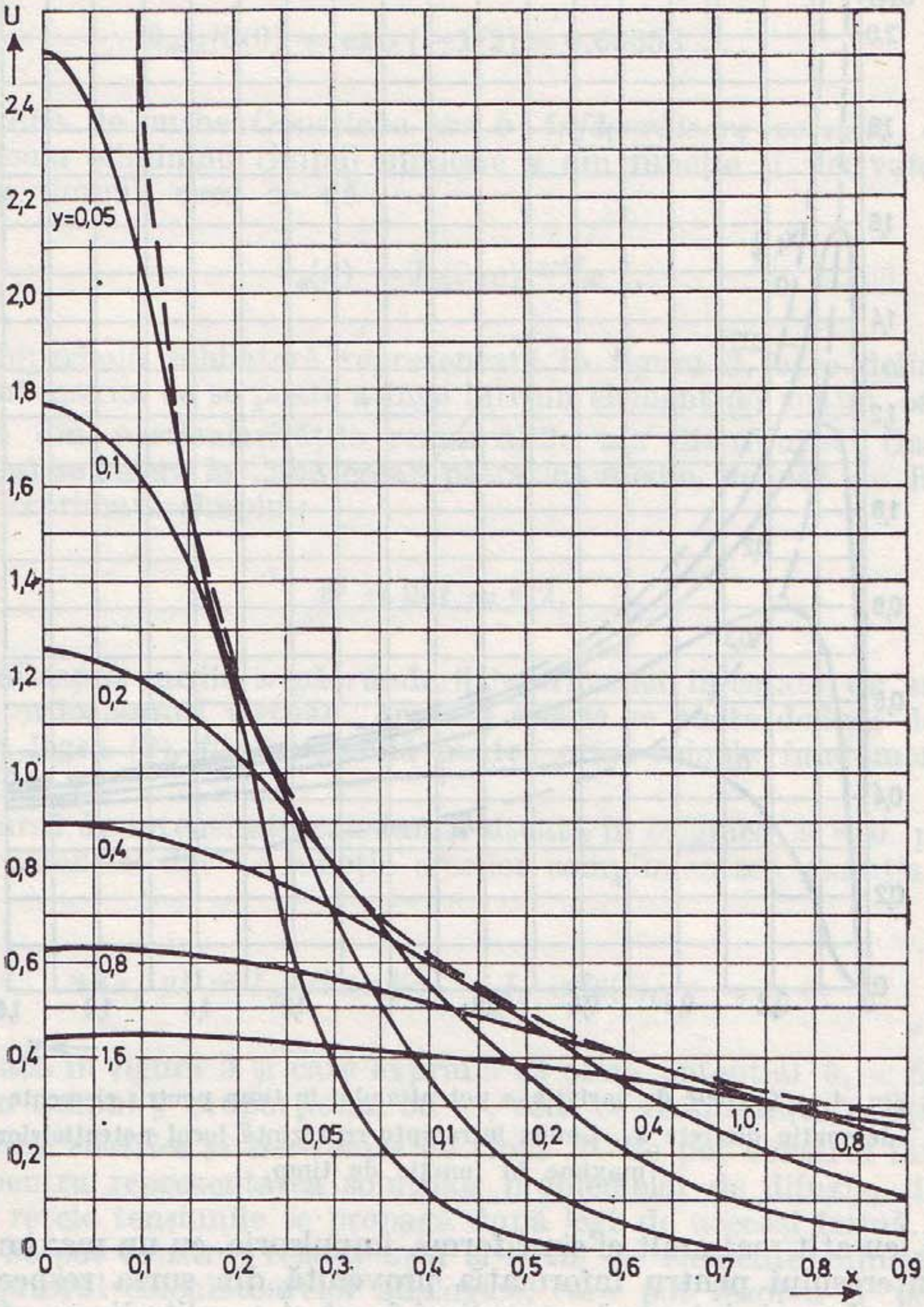

Fig. 3. — Familia de curbe clopot (Gauss) a distribuției erorilor de măsură aleatorii, soluție impulsorie a ecuației diferențiale Fourier (7). Liniile întrerupte reprezintă înfășurătoarea curbelor, locul potențialelor maxime în elementele de spațiu x_i.

deci el scade hiperbolic, invers proporţional cu radicalul din timp, comportare binecunoscută în teoria difuziei. În modelul informatic, interesul pentru informaţia sursei scade după această lege în timpul procesului de difuzie.

Considerînd spaţiul autorilor virtuali ca parametru şi înscriind timpul în abscisă, rezultă curbele de potenţial informaţional din figura 4, care permit o concluzie foarte preţioasă : cu cît autorii se află mai aproape

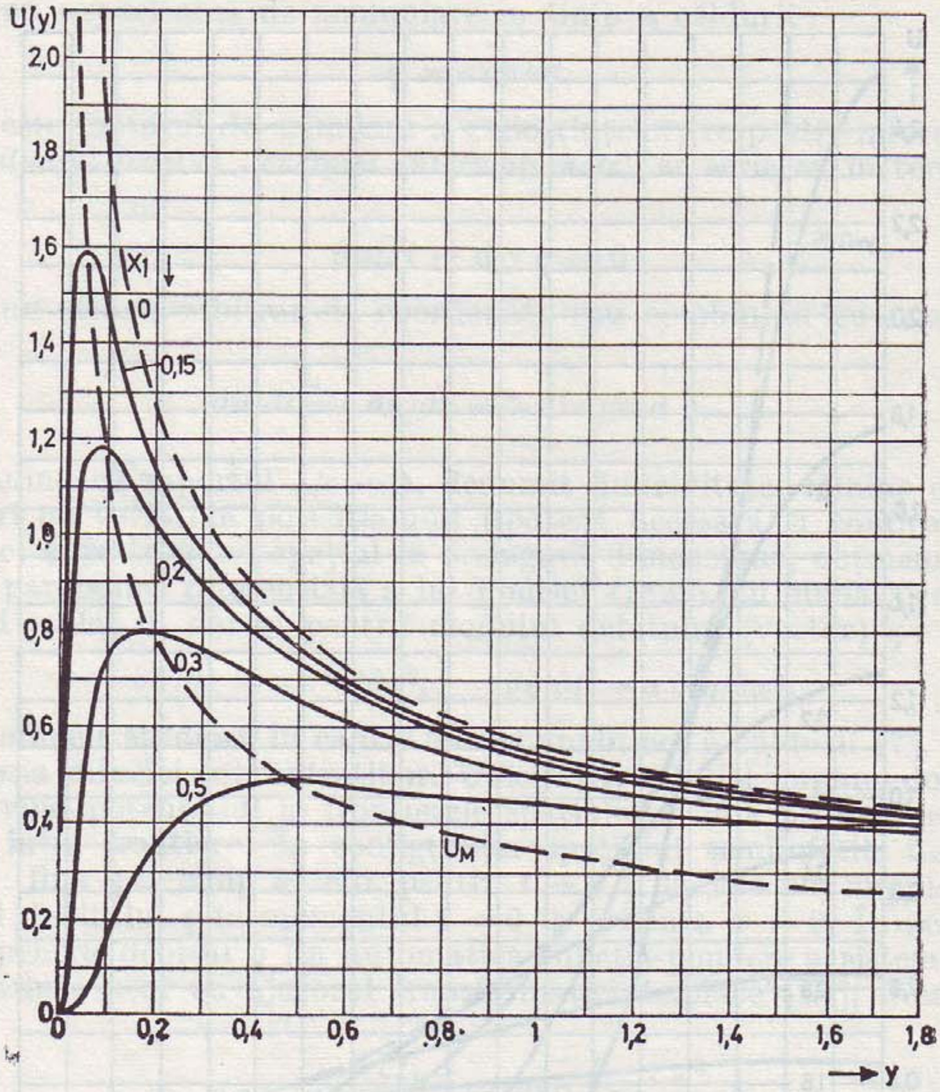

Fig. 4. — Curbele de variaţie a potenţialului în timp pentru elemente de spaţiu discrete x_i. Liniile întrerupte reprezintă locul potenţialelor maxime în funcţie de timp.

de sursă, cu atît mai mult ei simt forma impulsorie, cu un maxim pronunţat al interesului pentru informaţia provenită din sursa respectivă. La distanţe mai mari, interesul creşte mai încet şi amplitudinea, atinsă mai tîrziu, scade. De aceea, tinerii care doresc să devină cercetători se plasează instinctiv cît mai aproape de o sursă fecundă, de un maestru cu reputaţie, de la care pot să obţină informaţii cît mai „calde", de un cît mai mare interes.

Locul maximelor se obține anulînd derivata expresiei (9a), anume pentru timpuri $y = 2x^2$ și introducînd această relație în ecuația curbelor Gaussiene :

$$\vartheta_{\max} = I_0(\pi y)^{-1/2} \exp(-1/2). \tag{10}$$

Aceste maxime se află într-un raport constant față de potențialul sursei (9b), anume

$$\vartheta_{\max}/\vartheta(0) = \exp(-1/2) = 0{,}60653\ \ldots \qquad 10a)$$

Familia de curbe Gaussiene are o *înfășurătoare comună* care se poate calcula eliminînd timpul numeric y din funcție și derivata ei în raport cu timpul, ceea ce dă :

$$\vartheta_0(x) = I_0(2\pi e)^{-1/2}\, x^{-1}, \tag{11}$$

adică o hiperbolă echilateră reprezentată în figura 3, care delimitează potențialul maxim ce se poate atinge într-un element de spațiu oarecare.

Una din particularitățile remarcabile ale distribuției Gaussiene (9) sau (9a) se referă la „abaterea" pătratică medie, dedusă de Einstein în 1905, exprimată simplu :

$$\bar{x}^2 = 2at = y/2, \tag{12}$$

care dă distanța medie a pătrunderii informației injectate de sursă în masa cititorilor/autori virtuali. Această relație se poate deduce de altfel direct din legea (3), fiind valabilă pentru orice soluție fundamentală a ecuației (7).

O sursă de intensitate constantă situată în originea $x = 0$ produce un cîmp potențial dat de funcția erorilor complimentară (soluția „indicială") :

$$\vartheta_1(x, y) = I_0 \operatorname{erfc}(x/\sqrt{y}) = I_0 \operatorname{erfc}(z), \qquad z > 0, \tag{13}$$

reprezentată în figura 5 și care exprimă că orice potențial $\vartheta_i < \vartheta_0$ poate fi atins în timpul y proporțional cu x^2, ceea ce reamintește relația (12).

Modele electrice și electronice de tipul *RC* se pot construi fără dificultăți, pentru reprezentarea soluțiilor problemelor de difuzie, deoarece în aceste rețele tensiunile se propagă după legi de aceeași formă. Astfel de rețele se pot construi relativ ușor și ieftin cu elemente miniaturizate sau cu ajutorul calculatoarelor analogice, care pot rezolva și probleme neliniare.

Prima confirmare a teoriei difuziei s-a obținut tocmai prin utilizarea relației (12) în cazul unei surse permanente, caracterizînd un nivel constant al potențialului informațional în oricare element spațial, ceea ce

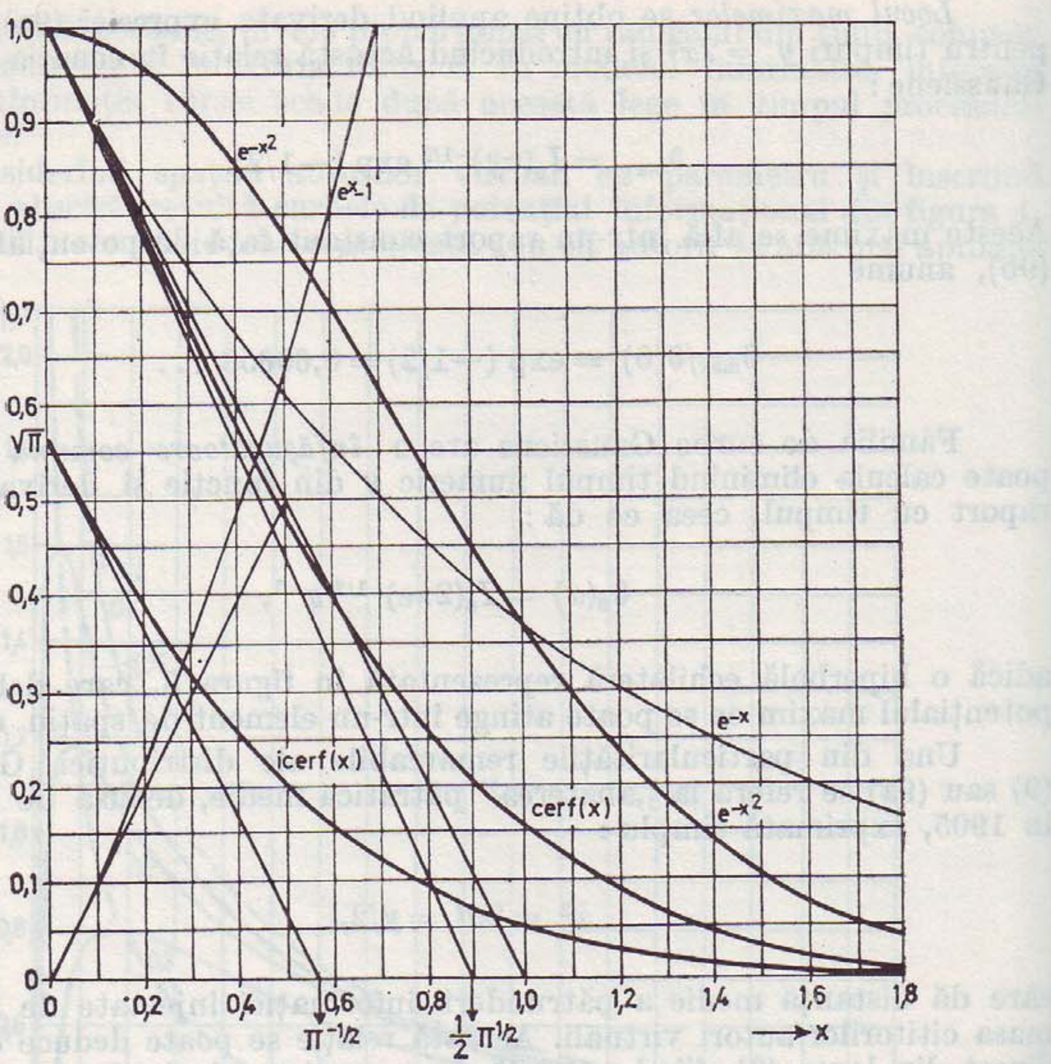

Fig. 5. — Funcţia complimentară şi integrală a erorilor (Gauss) comparată cu funcţia exponenţială simplă şi pătratică.

a contribuit la deducerea legii lui Bradford [8]. Această legitate bibliometrică, exprimată de autorul ei verbal, constată că numărul de reviste în care apare un număr dat de articole privind aceeaşi problemă creşte în progresie geometrică, aşa cum se vede pe porţiunea dreaptă ascendentă în figura 6. Relaţia analitică se deduce imediat, dacă se elimină timpul între inversa expresiei (1) a creşterii numărului de reviste şi relaţia (12). Cu notaţia $X = x^2/2$, reprezentînd numărul cumulativ al revistelor, rezultă

$$X(j) = (1/4)\ Y, \ln(j - j_0), \tag{14}$$

formal identică cu legea Bradford. Cum condiţia x^2/y = const. înseamnă potenţial constant, legea Bradford se bazează pe condiţia plauzibilă că lucrările privind acelaşi obiect de cercetare trebuie să aibă un nivel aproximativ egal.

Autorul a dedus însă din teoria difuziei nu numai acest aspect central al legii, ci și anomaliile ce apar la începutul și sfîrșitul curbelor experimentale care au forma (neexplicată) sigmoidă sau logistică, așa cum se vede în figura 6. În acest scop, se presupune în origine o sursă avînd un potențial crescător în timp de forma $\vartheta(0, t) = \vartheta_0(\pi t)^{1/2}$, pentru care soluția ecuației diferențiale (7) se scrie:

$$\vartheta(x, y) = \vartheta_0(ky)^{1/2} \ \mathrm{ierfc}(x/\sqrt{y}), \quad k = \pi/4a, \tag{15}$$

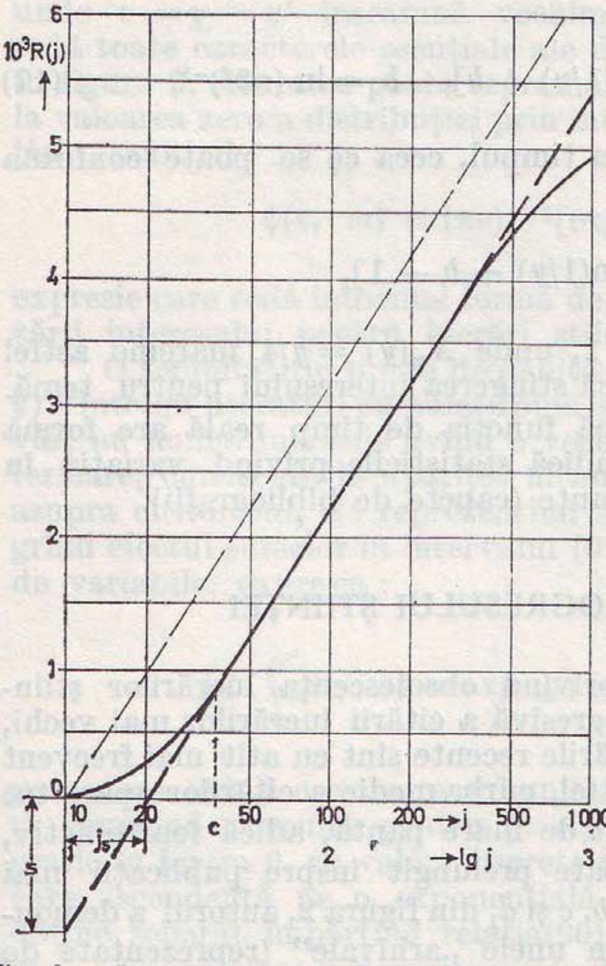

Fig. 6. — Legea distribuției Bradford a numărului cumulat de articole $X = R$, răspîndite în reviste j, confirmată prin numeroase bibliografii mari privind aceeași problemă.

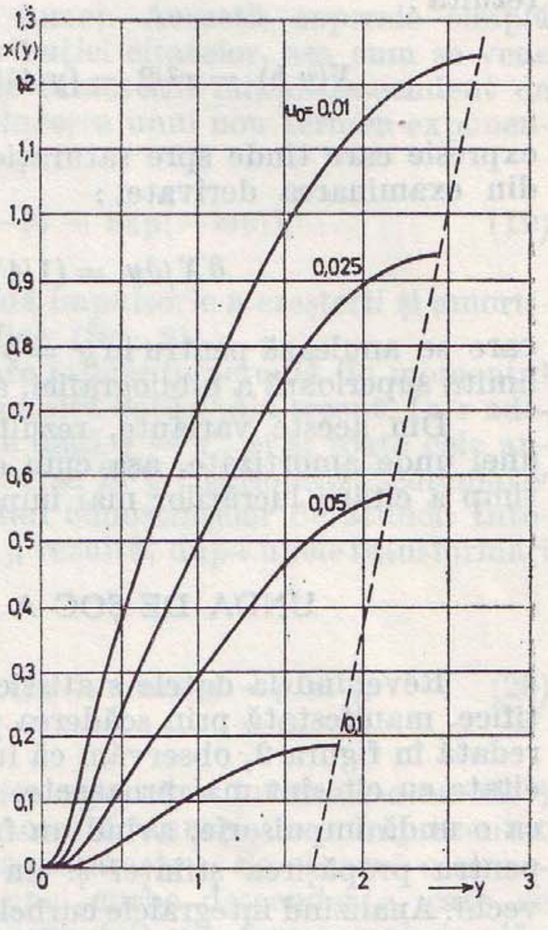

Fig. 7. — Curbe Bradford deduse din teoria difuziei pentru surse inițiale avînd o variație de creștere rapidă și apoi de scădere amortizată a potențialului de interes. Justificarea anomaliilor apărute în statistici. Linia întreruptă: limita superioară a bibliografiilor.

cu ajutorul integralei funcției complimentare a erorilor, care este tabelată [10], [13]. Astfel, pentru valori discrete, această funcție se poate inversa rezultînd:

$$X(y) = x^2/2 = (y/2) \ \mathrm{ierfc}^{-1} \ (\vartheta/\sqrt{ky}), \tag{16}$$

în funcție de timpul numeric y, care se mai poate elimina inversînd relația (1). Dealtfel, implicarea timpului în aceste formule se poate considera binevenită, deoarece bibliografiile se întind în timp într-un fel încă practic neinvestigat. În orice caz, aşa cum se poate vedea în figura 7, funcţia (16) reproduce în mod corespunzător comportarea bibliografiilor în apropiere de origine, adică în aşa-numitul „miez'' iniţial al revistelor care a servit lui Bradford ca bază de comparaţie cu grupările de reviste următoare. Din figura 7 mai rezultă că bibliografiile de înalt nivel sînt scurte şi invers, ceea ce pare evident.

Pe de altă parte, presupunînd nivele discrete de potenţial în cazul sursei impulsorii, tratate mai sus, din inversarea soluţiei (9a) rezultă :

$$X(y,b) = x^2/2 = (y/4)[\ln(1/y) + b], \quad b = \ln(\pi\vartheta^2)^{-1}, \tag{17}$$

expresie care tinde spre saturaţie cu timpul, ceea ce se poate confirma din examinarea derivatei :

$$\partial X/\partial y = (1/4)\,[\ln(1/y) + b - 1],$$

care se anulează pentru $\ln y = b - 1$, unde $X_m(y) = y/4$, marcînd astfel limita superioară a bibliografiei, adică stingerea interesului pentru temă.

Din aceste variante, rezultă că funcţia de timp reală are forma unei unde amortizate, aşa cum o indică statisticile privind variaţia în timp a citării lucrărilor mai importante (capete de bibliografii).

UNDA DE ŞOC A PROGRESULUI ŞTIINŢEI

Revenind la datele statistice privind obsolescenţa lucrărilor ştiinţifice, manifestată prin scăderea progresivă a citării lucrărilor mai vechi, redată în figura 2, observăm că lucrările recente sînt cu atît mai frecvent citate cu cît sînt mai proaspete. Astfel, curba medie a citărilor apare tot ca o undă impulsorie, avînd un front de mare pantă, adică foarte activ, pentru propăşirea ştiinţei şi un spate prelungit înspre publicaţii mai vechi. Analizînd integralele curbelor b, c şi d_1 din figura 2, autorul a demonstrat [6] că bipartiţia lucrărilor, în unele „arhivale'' (reprezentate de dreapta b) şi cele active (exponenţiala c), nu se poate justifica logic, deoarece nu dispunem de nici un criteriu de selecţie valabil. Autorul a căutat să deducă din teoria difuziei o expresie analitică pentru curba cumulativă d, care să corespundă în întregul domeniu cu alura dată de datele statistice. Această tentativă a reuşit, premisele şi rezultatele fiind prezentate pe scurt în cele ce urmează.

Interesul pentru o lucrare anterioară, apărută la timpul $y' > 0$ şi resimţit în momentul $y > y'$, se compune din efectul de obsolescenţă (9b) al unei surse şi creşterea exponenţială (2) a surselor succesive, scriindu-se :

$$\vartheta(y) = \vartheta_0[\pi(y - y')]^{-1/2}\exp(y'), \quad \vartheta_0 = 2I_0X_0Y_x^{-1/2}, \tag{18}$$

în care constanta ϑ_0 depinde de valorile inițiale (la $y = 0$) și de constanta de timp Y_x a creșterii citatelor. Dealtfel constanta ϑ_0 fiind irelevantă pentru funcția de timp a potențialului, se poate utiliza ca referință, împreună cu numărul surselor la timp $y > y'$, rezultînd :

$$\varphi(y, y') = (\vartheta/\vartheta_0) \exp(-y) = [\pi(y - y')]^{-1/2} \exp(y' - y) = \\ = (\pi v)^{-1/2} \exp(-v), \tag{18a}$$

unde $v = y - y'$ înseamnă vechimea sursei. Această expresie simplă redă toate caracterele esențiale ale distribuției citatelor, așa cum se vede în figura 2. Acum se poate reprezenta și caracterul inițial descendent de la valoarea zero a distribuției prin introducerea unui nou termen exponențial, rezultînd :

$$\psi(v, m) = (\pi v)^{-1/2}[\exp(-v) - \exp(-mv)], \tag{19}$$

expresie care redă întocmai forma de undă impulsorie a creșterii și amortizării interesului pentru lucrări științifice (fig. 8).

O semnificație și mai deosebită o are influența actuală (în momentul y) a întregii literaturi cumulate pînă la anumită dată y' din trecut. Într-adevăr, nu numai lucrările avînd o vechime dată $y - y'$, ci și toate cele anterioare, cunoscute la apariția ultimei surse în y', acționează cumulativ asupra cititorului, ele reprezentînd stadiul cunoștințelor de atunci. Integrînd efectul surselor în intervalul $[0, y']$, rezultă, după unele transformări de variabile, expresia :

$$\vartheta_0^{-1} \int_0^{y'} \vartheta(y)\, \mathrm{d}y = \exp(y)\, [\mathrm{erf}\, (y)^{1/2} - \mathrm{erf}\, (y - y')^{1/2}], \tag{20}$$

paranteza avînd o valoare subunitară (ca a oricărei integrale a erorilor) și devenind maximă pentru $y = y'$. Această funcție s-a reprezentat grafic în figura 9, cu valori discrete y' ca parametru. Se observă o alunecare ascendentă pe o exponențială a unei curbe descendente, care se obține separat împărțind relația(20) cu $\exp(y')$ sub forma mai simplă :

$$\frac{\theta(y)}{\vartheta_0 \exp(y')} = \exp(v)\, \mathrm{erfc}\,(v), \qquad v = y - y', \tag{21}$$

dacă totodată presupunem $y > 6$, ceea ce revine la a considera o perioadă de cel puțin 6 constante de timp, adică circa 100 ani. Această funcție depinde numai de vechimea v, deoarece $\mathrm{erf}(2{,}5) \to 1$ (tinde spre unitate) și paranteza se poate exprima prin complimentara integralei erorilor, funcție eminamente descrescătoare pentru argumente reale pozitive. Acum creșterea exponențială a surselor se poate considera prin trecerea termenului $\exp(y')$ în membrul drept. În concluzie, constatăm că efectul obsolescenței prevalează față de cel al creșterii numărului de lucrări, respectiv de citate.

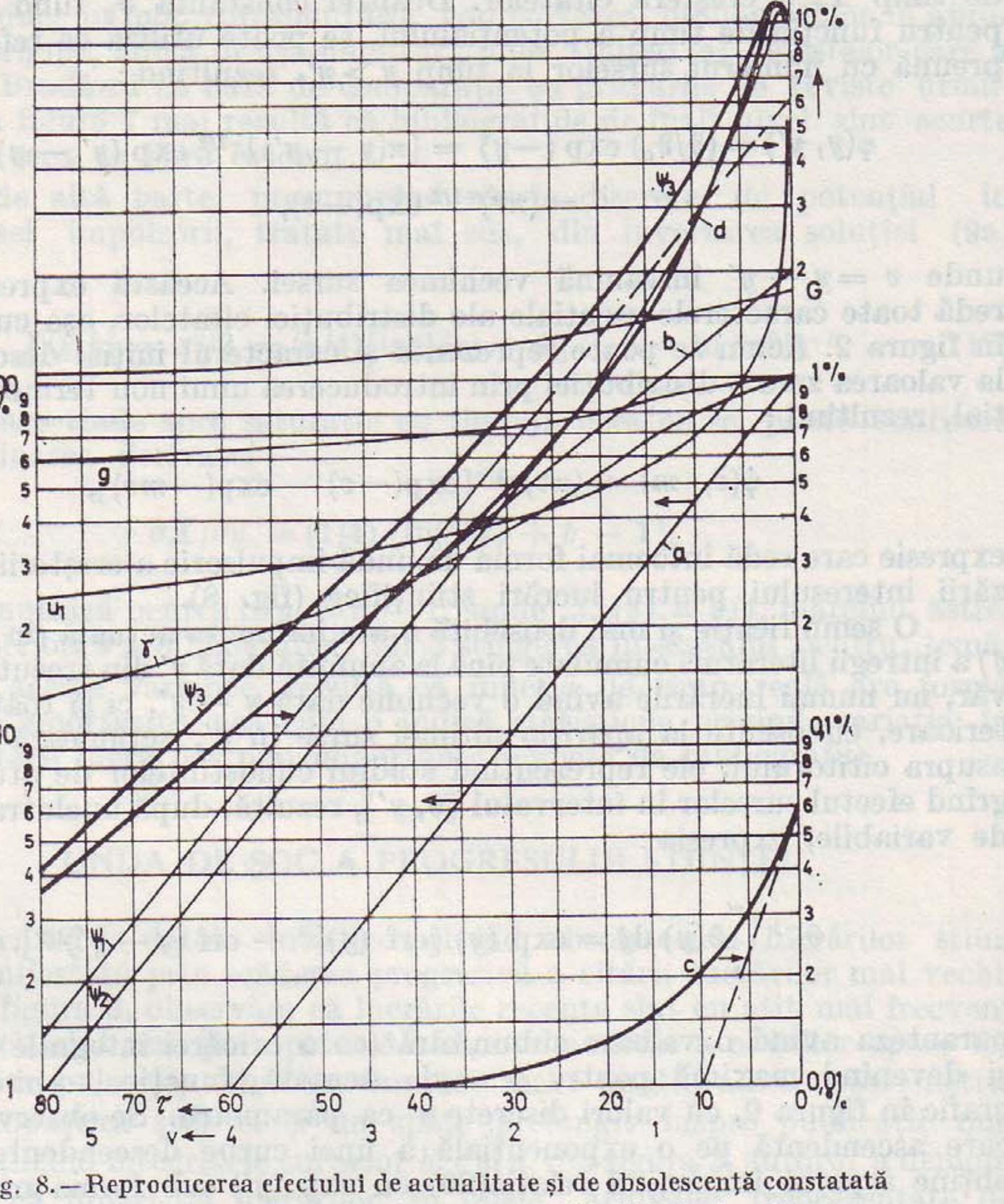

Fig. 8. — Reproducerea efectului de actualitate și de obsolescență constatată a articolelor științifice prin soluția temporală (19) în funcție de vechimea $v = y - y'$ a surselor de informație:

u_1 — obsolescența locală a surselor;
d — obsolescența constatată a referințelor (citatelor);
G — rata de scădere a numărului citărilor multiple;
γ — rata de scădere relatată a citărilor multiple;
ψ_1 — unda de interes dedusă ($m = 10$, $T = 15$ ani);
ψ_2 — idem ($m = 20$, $T = 15$ ani);
ψ_3 — idem ($m = 20$, $T = 20$ ani), aproximație ameliorată.

Distribuția în spațiul x se poate lua în considerare cu un factor suplimentar în soluția (20):

$$\theta(x, y, y') = \vartheta_0 \exp(y) \operatorname{erfc}(y - y')^{1/2} \exp\left[-\frac{(x - x')^2}{y}\right], \quad (22)$$

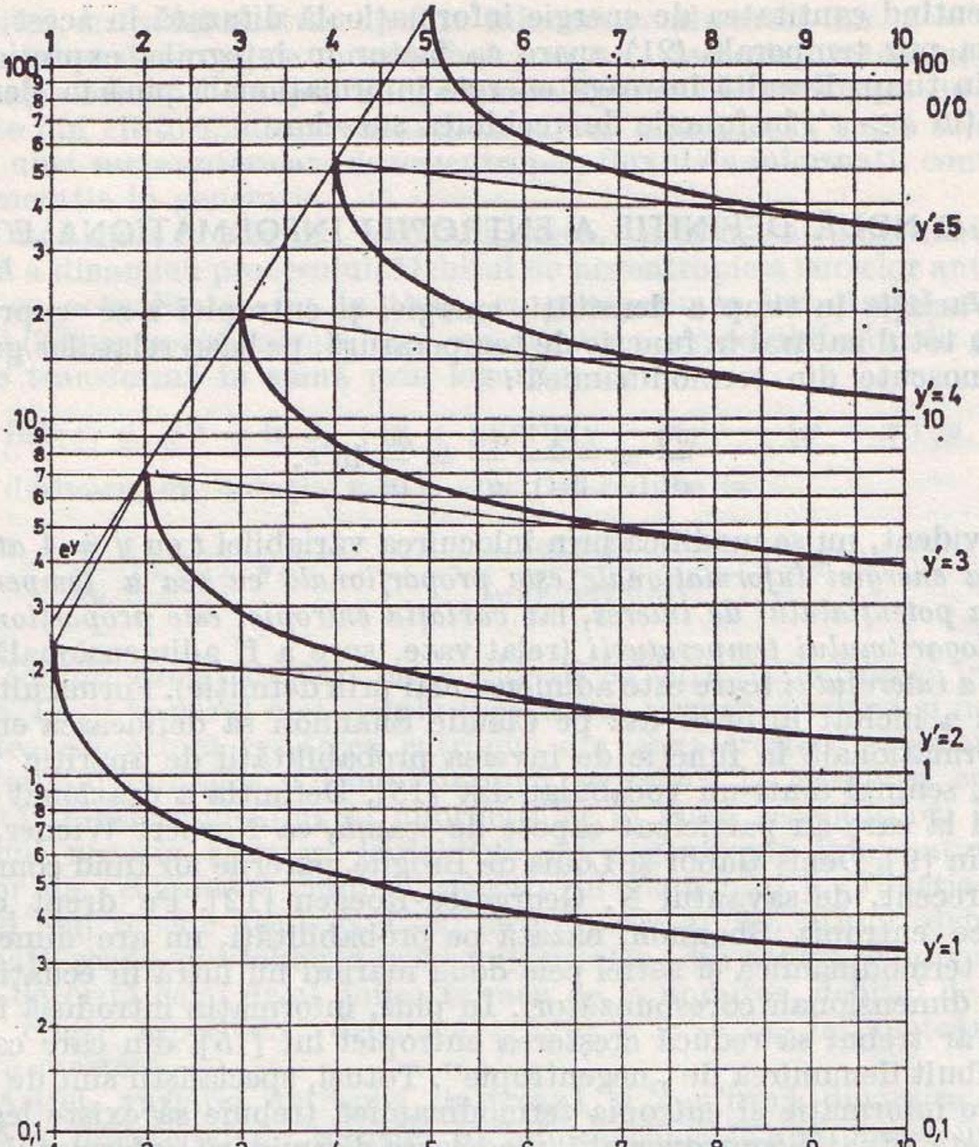

Fig. 9. — Efectul cumulat al surselor mai vechi decît y' în momentul $y > y'$. Se observă creșterea exponențială a surselor și scăderea bruscă a interesului imediat după apariția ultimei surse în y'. Mai tîrziu interesul scade mult mai lent cu timpul.

care exprimă potențialul de interes în masa $x > x'$ a cititorilor/autori virtuali, față de surse anterioare mai vechi decît y'.

În orice caz, aceste funcții exprimă și reproduc calitativ unda distribuției de citări, mărturie a legăturilor strînse între lucrări originale vechi și noi, astfel cum ele au fost constatate empiric, ceea ce se poate vedea din comparația efectuată în figura 9.

Integrarea în spațiul $x > 0$ a ultimului factor din soluția (22) dă:

$$\int_0^\infty \exp\,[-(x-x')^2/y]\,\mathrm{d}x = (1/2)\sqrt{\pi y}\ \{\mathrm{erf}[(x-x')/\sqrt{y}] - \mathrm{erf}(x'/\sqrt{y})\}, \quad (23)$$

reprezentînd cantitatea de energie informaţioală difuzată în acest spaţiu. Funcţia pur temporală (21) apare ca factor în integrala expresiei (22) luată în timp. Rezultă întreaga energie informaţională pînă în elementul de spaţiu $x > x'$, în funcţie de vechimea surselor.

O NOUĂ DEFINIŢIE A ENTROPIEI INFORMAŢIONALE

Variaţia în timp a densităţii energiei şi entropiei s se exprimă în mod cu totul natural în funcţie de temperatură, pe baza relaţiilor generale binecunoscute din termodinamică :

$$\frac{\partial s}{\partial t} = \frac{1}{\vartheta(t)} \frac{\partial u}{\partial t} = \frac{c\partial}{\partial t} \ln \vartheta, \tag{24}$$

care, evident, nu se modifică prin înlocuirea variabilei t cu $y = 4\ at$. Deci, *variaţia energiei informaţionale este proporţională cu cea a temperaturii*, adică *a potenţialului de interes*, iar *variaţia entropiei este proporţională cu cea a logaritmului temperaturii* (relativate, spre a fi adimensională), respectiv *a interesului* (care este adimensional prin definiţie). Forma ultimului termen a incitat în orice caz pe Claude Shannon să definească entropia sa informaţională în funcţie de inversa probabilităţii de apariţie a unui anumit semnal dintr-un vocabular dat [15]. Definiţia a ocazionat ample discuţii la care au participat capete de seamă, ca Norbert Wiener, Léon Brillouin [9], Denis Gabor şi Louis de Broglie, părerile lor fiind comparate critic, recent, de savantul N. Georgescu-Roegen [12]. Pe drept cuvînt, deoarece entropia Shannon, bazată pe probabilităţi, nu are dimensiune ca cea termodinamică şi astfel cele două mărimi nu intră în ecuaţii, fără factori dimensionali corespunzători. În plus, informaţia introdusă într-un sistem ar trebui să reducă creşterea entropiei lui [15], din care cauză i s-a atribuit denumirea de „negentropie". Totuşi, specialiştii sînt de părere că între informaţie şi entropia termodinamică trebuie să existe legături. În această situaţie controversată, modelarea diseminării informaţiilor prin fenomene fizice de difuzie ne îndreptăţeşte să utilizăm fără reticenţe relaţiile din termodinamică, interpretîndu-le în mod corespunzător în informaţia documentară. În cazul entropiei, beneficiul modelării se dovedeşte a fi considerabil.

Trecînd în ecuaţiile (24) la timpul numeric y, ceea ce nu le modifică forma şi introducînd soluţia Gaussiană pentru temperatură, conform ecuaţiei (9a) găsim :

$$\ln \vartheta\,(x, y) = -x^2/y - (\ln \pi + \ln y)/2 \tag{25}$$

şi derivata ei :

$$\frac{\partial}{\partial y} \ln \vartheta\,(x, y) = x^2/y^2 - 1/2y = (1/c)\ \partial s/\partial y. \tag{26}$$

Deci variaţia entropiei conţine doi termeni, unul pozitiv, de entropie crescătoare cu pătratul distanţei şi descrescătoare cu pătratul timpului, şi unul negativ, de negentropie crescătoare invers proporţional cu timpul. Primul termen se referă la disipaţia în spaţiul $x > 0$, iar al doilea la debitul, în cazul nostru, de flux informativ al sursei din $x = 0$. Cele două mărimi devin egale numai cînd $y = 2x^2$, deci cînd interesul pentru informația oferită de sursă atinge un maximum în elementele de spațiu, formate din cititori/autori virtuali. Scăderea entropiei în sursă se face pe seama unei surse anterioare de negentropie, fluxul de informații continuînd din generație în generație.

După cum se știe din termodinamică, variația entropiei mai este o măsură a dinamicii procesului. Debitul de negentropie a surselor anterioare crește cu actualitatea lor în timp, cu prospețimea lor. Astfel, entropia soluției (22) se poate calcula ușor, ca urmare a aspectului său de produs, care se transformă în sumă prin logaritmare:

$$\ln \theta(x, y, y') = \ln \vartheta_0 + y + \ln \operatorname{erfc}(y - y')^{1/2} - (x - x')^2/y. \quad (27)$$

După derivare, cu notația $v = y - y'$, se reduce la

$$\frac{\partial}{\partial y} \ln \theta\,(x, y, y') = 1 - \frac{\exp(v)}{(\pi v)^{1/2}\operatorname{erfc}(v)} + (x - x')^2/y, \quad (28)$$

fiind, conform relației (24), proporțională cu variația entropiei informaționale. Ca și în cazul mai simplu (26), avem termeni pozitivi și negativi. Unitatea reprezintă creșterea entropiei din cauza creșterii exponențiale a potențialului citatelor. Ultimul termen se referă la creșterea entropiei datorită transferului difuz al informațiilor în masa receptorilor. În fine, termenul negativ se referă la debitul de negentropie informațională a surselor de vechime v, resimțit selectiv în momentul y. Funcția $\operatorname{erfc}(z)$ scăzînd mai repede decît $\exp(-z)$ la creșterea argumentului, debitul de informații scade cu vechimea frontului surselor; din contra, cînd vechimea scade, datorită celui de-al doilea termen de la numitor, debitul de negentropie crește hiperbolic, deoarece $\operatorname{erfc}(0) = 1$ și nu influențează deci această creștere.

Astfel, variația entropiei ilustrează și confirmă dinamica excepțională a procesului de transfer al informațiilor în apropierea imediată a frontului de surse, așa cum ea rezultă din constatările statistice. De remarcat că în relațiile (24) nu intervine valoarea absolută a entropiei, ci numai variația ei în timp, dar tocmai aceasta interesează mai mult. Într-adevăr, din relația de definiție (24) reiese că variația în timp a densității energiei, deci a puterii locale (5), rezultă din produsul temperaturii și al densității de entropie, pe acest fapt bazîndu-se grafurile utilizate recent (bond graphs) pentru studiul termodinamic al mașinilor de forță [18]. Fluxul de entropie este strîns legat de transportul de energie termică, respectiv informațională, la fel ca și mișcarea sarcinilor (curentul) de transportul energiei electrice. În procesul tipic ireversibil al conducției căldurii în solide și, similar, în acel al difuziei informațiilor, odată cu transferul fluxului prin elementele spațiului se generează entropie, cu toate că fluxul se conservă. Temperatura, respectiv interesul pentru lucrări anterioare, descrește, dar fluxul de entropie crește. Raportul dintre potențial și fluxul de entropie, care s-ar putea denumi ,,impedanță'', trebuie să descrească în orice proces ireversibil.

Încă o analogie se descoperă la nivelul temperaturii absolute din termodinamică, căreia în informatica bibliometrică îi corespunde potenţialul de fond al cunoştinţelor actuale, la care se adaugă contribuţii suplimentare prin lucrări originale nou-apărute. În fond, ne interesează mai mult rata creşterii acestor suplimente cît mai recente, din care putem să deducem ritmul procesului de dezvoltare a ştiinţei.

În orice caz, definiţia (24) a entropiei termodinamice se poate aplica în acelaşi sens şi fără restricţii la procesul de difuzie a informaţiilor ştiinţifice în spaţiul cititorilor de articole şi viitorilor autori, contribuind la ordonarea cunoştinţelor, la extinderea lor în direcţiile preferenţiale şi actuale ale cercetărilor mondiale. Reproducerea şi lămurirea unor legităţi bibliometrice şi deducerea multor relaţii ascunse între parametri măsurabili confirmă valabilitatea modelului difuziv propus pentru caracterizarea procesului de diseminare a informaţiilor ştiinţifice. De asemenea, definirea lipsită de ambiguităţi a entropiei informaţiei şi asimilarea, de data aceasta reuşită, a negentropiei cu variaţia debitului de informaţie a surselor aduc elemente noi în controversata problemă a legăturilor dintre entropia termodinamică şi informaţia introdusă într-un sistem izolat. Dealtfel, generalitatea termodinamicii şi marea putere a teoriei analitice a căldurii oferă, în cadrul modelării prezentate, largi posibilităţi de dezvoltare a teoriei difuziei informaţiilor ştiinţifice, amplificate încă dacă s-ar putea realiza o legătură corespunzătoare cu teoria modernă a sistemelor şi cea a telecomunicaţiilor.

BIBLIOGRAFIE

1. AVRAMESCU A., *Möglichkeiten einer Wertzuordnung an wissenschaftliche Veröffentlichugen*, Dokumentation und Information. XI Internat. Kolloquium T.H., Ilmenau, R.D.G., oct. 1966, 35–47.
2. AVRAMESCU A., Stud. Cerc. Docum., 1973, **15**, *1*, 3–19.
3. AVRAMESCU A., Stud. Cerc. Docum., 1973, **15**, *3*, 243–253.
4. AVRAMESCU A., Stud. Cerc. Docum., 1973, **15**, *4*, 345–356.
5. AVRAMESCU A., Stud. Cerc. Docum., 1973, **17**, *1–2*, 25–49.
6. AVRAMESCU A., Internat. Forum on Inform. and Docum. Moscow, 1976, **1**, *1*, 13–19.
7. BECKER R., *Theorie der Wärme*, Springer Verlag, Berlin-Göttingen-Heidelberg, 1964.
8. BRADFORD S. C., Engineering, London, 1934, *137*, 85–88.
9. BRILLOUIN L., *Science and Information Theory*, Academic Press, New York, 1963.
10. CARSLAW H. S., JAEGER J. C., *Conduction of Heat in Solids*, Clarendon Press, Oxford, 1962.
11. GARFIELD E., SHER I. H., Amer. Docum., 1963, **14**, 195–204.
12. GEORGESCU -ROEGEN N., *Entropy and the Economic Process*. App. B. *Ignorance, Information and Entropy*, Harvard Univ. Press, 1975.
13. JAHNKE-EMDE-LÖSCH, *Tables of Higher Functions*, B. G. Teubner, Stuttgart, 1960.
14. LOTKA A., J. Wash Acad. Sci., 1926, **16**, *12*, 317–323.
15. PETERS J., *Einführung in die allgemeine Informationstheorie*, Springer Verlag, Berlin-Heidelberg-New York, 1967.
16. PRICE D. de SOLLA, *Little Science Big Science*, Columbia Univ. Press, New York-London, 1963.
17. PRICE D. de SOLLA, *Networks on Scientific Papers*, în *The Growth of Knowledge*, New York-London-Sidney, 1967, 510.
18. THOMA J. U., J. Franklin Inst., 1971, **292**, 109–120.
19. TRAPEZNIKOV V. A., *Control, Economy, Technological Progress*, Proc. III IFAC Congress, London, Elsevier, 1966.

Academia Republicii Socialiste România
Secţia de ştiinţe tehnice
Calea Victoriei, nr. 125
Bucureşti