



UNIVERSITATEA BABEȘ-BOLYAI

REZUMATUL TEZEI DE DOCTORAT

Universalități statistice în sisteme socio-economice

Autor:
GERE István

Conducător de doctorat:
Prof. Dr. NÉDA Zoltán

*Rezumatul tezei
pentru obținerea
titlului de
Doctor în Fizică*

Universitatea Babeș-Bolyai
Cluj-Napoca, România
Februarie 2022

Universalități statistice în sisteme socio-economice

ABSTRACT

Analizând datele pentru procese socio-economice au permis identificarea unor universalități statistice și o descriere teoretică a acestora prin modele împrumutate din fizica statistică. Distribuția veniturilor și a averii au fost investigate și modelate în societățile contemporane și trecute. Pentru Internet, ca rețea fizică, au fost observate și modelate legi de scalare.

Folosind datele exhaustive pe mai mulți ani privind venitul în județul Cluj s-a observat o scalare în funcția de densitate a probabilității. Urmărind dinamica veniturilor personale, un model evolutiv (modelul de Creștere Locală cu Resetarea Globală, LGGR) a fost propus pentru o descriere reală a statisticilor observate. Universalitatea observată pentru funcția de densitate de probabilitate a fost verificată și pentru alte țări.

O abordare similară a fost aplicată pentru a descrie funcția de distribuție a averii în societățile contemporane. Metoda noastră a permis o descriere unificată a funcției de densitate de probabilitate pentru toate categoriile de avere, precum și descrierea regimului de avere negativă (datorie). Pentru multe țări cu diferite caracteristici economice a fost aplicată cu succes metoda de modelare LGGR a funcției de distribuție a averii.

Pentru o înțelegere mai completă a funcției de distribuție a averii, au fost analizate registrele agricole și evidențele fiscale pentru a estima distribuția averii într-o zonă delimitată geografic, comuna Sâncraiu (județul Cluj, România). Au fost analizate și modelate diferite perioade ale istoriei economice a comunei. S-a observat o funcție de distribuție a averii similară cu cea Gaussiană în perioada comunistă și o funcție de distribuție Tsallis-Pareto a averii pentru anul 2021 (tipic pentru societățile moderne), la 32 de ani de la desființarea regimului comunist în România. Abordarea LGGR a permis o bună descriere a funcției de distribuție a averii în fiecare perioadă, cu ratele de creștere și resetare care sunt în concordanță cu politicile economice din perioadele investigate. Indicatoarele de inegalități socio-economice au fost studiate din datele experimentale și au fost comparate cu rezultatele modelului.

Ultimul studiu prezentat în teză tratează evoluția topologiei Internetului. Au fost identificate legi noi de scalare. S-a propus un model simplu și realist din punct de vedere economic care generează grafuri cu proprietăți similare Internetului la nivel de router. Acest model reproduce distribuția conexiunilor observate în experimente simple. Modelul reproduce și scalarea observată pentru numărul de routere intermediare (în rețea fizică de Internet) în funcție de distanța geodezică dintre acestea.

În această teză se ilustrează prin exemple concrete utilitatea modelelor simple de fizică statistică în înțelegerea universalităților unor fenomene socio-economice complexe.

Cuvintele cheie: universalități statistice, econo-fizică, socio-fizică, sisteme complexe, fizică statistică, distribuție a venitului, distribuție a averii, model creștere locală și de resetare globală, metode de estimare a averii, legi de scalare, rețele, Internet, modelare bazată pe agenți

Cuprinsul tezei de doctorat

Abstract	i
Acknowledgements	ii
1 Introduction	1
1.1 The scientific motivation of the thesis	1
1.1.1 The Gaussian distribution	2
1.1.2 The exponential distribution	3
1.1.3 Power-laws, scaling and complexity	4
Detection	5
Complexity and power-laws	6
2 Statistical universalities for income and wealth distributions	8
2.1 Overview	8
2.1.1 Income versus wealth	9
2.1.2 Data gathering	10
2.2 Modelling the distributions of wealth and income	10
2.2.1 Model categories	12
2.3 The master equation approach, local growth and global reset model	12
2.3.1 Application of the LGGR for income distribution	14
Estimating the growth rate	15
The Reset	16
The probability density of income distribution	18
2.3.2 Results and discussions for the income distributions	20
A closer look over the income dynamics	23
2.3.3 Application to wealth distribution	25
Observations on the distribution	25
The growth and reset rate for wealth dynamics	28
Comparison with the experimental results	31
Remarks on the stationary distribution	33
Validity of the Smart-reset rate	34
Comparison with the Mezard & Bouchaud approach	35
2.4 Conclusion and discussion	38

3	A case study for wealth estimation and modelling	41
3.1	Overview	41
3.2	Exhaustive wealth data from Agricultural records and taxation databases	42
3.3	Wealth estimation methodology	43
3.3.1	Estimated wealth for 1961	44
3.3.2	Estimated wealth for 1989	46
3.3.3	Estimated wealth for 2021	48
3.4	Theoretical approach for the observed distributions	50
3.4.1	Private wealth under the communist regime? - A constant growth rate and a linear reset rate	51
3.4.2	The rich gets richer dynamics - Linearly increasing growth and a constant reset rate.	52
3.4.3	Validation of the theoretically obtained probability densities by the experimental wealth data	52
3.5	Dynamics in the inequality	54
3.6	Discussion	56
3.6.1	The observed wealth distributions	56
3.6.2	The LGGR approximation and the justification of the rates	57
3.6.3	Economic policy and inequality	59
3.7	Conclusion and outlook	59
4	Modelling the wiring topology of the Internet	61
4.1	Overview	61
4.1.1	Graphs, Networks and the Internet	61
	Graphs	61
	Network	63
	The Internet	63
	About the structure of the Internet	63
4.1.2	Relevant topological and dynamical properties of the Internet	64
	The scaling of the Round-Trip-Time and the geodesic distance	64
	The graph representation of the Internet as revealed by the Traceroute experiment	66
4.2	Modelling the Internet topology	67
	The properties of the generated graph	69
4.3	Scaling for the information propagation process	72
4.4	Conclusion and outlook	73
5	Conclusion	75
	List of Publications	78
	Bibliography	80

Cuprinsul rezumatului

Abstract	i
1 Introducere	1
2 Universalități statistice pentru distribuția veniturilor și a averilor	3
2.1 Abordarea ecuației master, Creșterea Locală și Resetare Globală	4
2.1.1 Aplicarea metodei LGGR pentru descrierea funcției de distribuție a venitului	5
2.1.2 Aplicarea metodei LGGR pentru descrierea funcției de distribuție a averii	6
2.2 Concluzie și discuție	8
3 Studiu de caz pentru estimarea averii și modelarea acesteia	10
3.1 Metodologia de estimare a averii	11
3.2 Abordare teoretică pentru descrierea distribuțiilor observate	12
3.3 Inegalități observate	13
3.4 Concluzie și discuție	14
4 Modelarea topologiei fizice de interconectare în rețeaua de Internet	16
4.1 Proprietăți dinamice și topologice relevante ale Internetului	16
4.2 Modelarea topologiei de rețea a Internetului	17
4.3 Scalare în procesul de propagare a informațiilor	19
4.4 Concluzie	19
5 Concluzie	21
Lista de publicații	23
Bibliografie selectată	24

Introducere

Analizarea datelor mari cu ajutorul calculatoarelor performante de astăzi, permite studierea universalităților statistice în multe domenii de studii interdisciplinare [1]. Aceste observații statistice pot fi înțelese cu metodele fizicii statistice.

Una dintre aceste universalități statistice fundamentale a fost descoperită înaintea erei calculatoarelor. Renumita **distribuție Gaussiană** este observată în multe fenomene (aparent fără legătură), de exemplu în distribuția circumferinței toracice a soldaților [2], sau la distribuția timpilor scurși între atacurile cibernetice executate împotriva serverului de fișiere aparținând departamentului nostru. Teorema limitei centrale declară că dacă o variabilă aleatoare este suma mai multor variabile aleatoare independente, atunci distribuția sumei va fi una Gaussiană [3]. Din acest punct de vedere, natura universală a distribuției gaussiene este clară, iar fenomenele influențate de factori aleatori independenți vor urmări o astfel de distribuție.

Un alt exemplu de universalitate statistică este ce a **distribuției exponențiale** care este caracteristică proceselor fără memorie. Aceasta este folosită de exemplu pentru a descrie timpii de așteptare pentru evenimente care au loc cu o rată constantă. De exemplu, timpul până la defectarea unei componente dintr-un dispozitiv [3] sau distribuția timpului măsurat până la descompunerea unui singur atom radioactiv [4].

În centrul studiilor prezentate în această teză se află **legile de putere** (power-laws). Multe fenomene complexe [11] produc distribuții care converg către o lege de putere (Eq. 1.1).

$$\rho_{PL}(x) \propto x^{-a} \quad (1.1)$$

Exemple de fenomene notabile care prezintă astfel de distribuții: distribuția de citări al articolelor științifice sau distribuția de "share"-uri al postărilor de pe Facebook [5], distribuția energiei eliberate de cutremure (legea Gutenberg și Richter) [6], distribuția populației orașelor [7], vitezele de deplasare pe rețelele de transport [8], sau chiar observația renumită a lui Pareto legată de distribuția averii [9].

O distribuție care urmărește o lege de putere pură nu poate fi normalizată pe intervalul $[0, \infty)$. Din această cauză distribuțiile observate experimental sunt de obicei descrise cu funcții care converg către o lege de putere. Exemple pentru astfel de funcții sunt: distribuția Beta-Prime, distribuția Tsallis-Pareto Type 2. (Lomax) etc.

Distribuțiile care urmăresc legi de putere sunt detectate prin analiza datelor referitoare la fenomenele observate, folosind metoda histogramelor, sau cea a curbelor cumulative. În studiile sale, Pareto a folosit curbele cumulative pentru avere [9].

Scalari și funcțiile de forma a legilor de putere sunt frecvente în fizica statistică. În cazul fenomenelor critice care au loc în jurul tranzițiilor de fază, se observă de obicei o scalare caracteristică, atunci când o anumită cantitate relevantă a sistemului diverge ca o funcție de putere a unei alte cantități relevante [10].

Din observații se pare că legile de putere sunt o proprietate intrinsecă a sistemelor complexe. Acest lucru este bine exemplificat și prin fenomene de Criticalitate Auto-Organizată (o stare dinamică statistic stabilă spre care tind multe sisteme) [11]. Prin această perspectivă, multe sisteme socio-economice pot fi considerate ca **sisteme complexe** [12].

Datele privind sistemele sociale și economice sunt constant colectate și arhivate, de către stat și de instituții private. Aceste date permit studiul și analiza necesară pentru înțelegerea proceselor care ghidează aceste sisteme. În această teză ne propunem să investigăm generalitatea distribuțiilor de tip legi de putere în sisteme socio-economice, confirmând universalitatea acestora prin modele și experimente.

În al doilea și al treilea capitol al acestei teze descriem studii privind distribuția veniturilor și averilor în societăți umane bine delimitate [13,14]. Se va descrie o metodologie simplă pentru estimarea averii în comunități mici și omogene și se vor explica distribuțiile observate experimental folosind un model propus recent, bazat pe ecuația master LGGR [15].

Al patrulea capitol este legat de scalarea proprietăților Internetului ca rețea fizică [16]. Propunem un model de rețea simplu care explică legile de scalare observate în experimente.

Universalității statistice pentru distribuția veniturilor și a averilor

În centrul multor studii socio-economice se află venitul și averea, deoarece acestea afectează viețile noastre, la nivel de individ, și la nivel de societate în ansamblu.

Încă de la descoperirea renumită a lui Vilfredo Pareto, la sfârșitul secolului al XIX-lea este cunoscut faptul că distribuția cumulativă a venitului și averii sunt legi de putere (Eq. 2.1.) [9].

$$P(X > x) \sim x^{-a} \quad (2.1)$$

Distribuția averii ca și o lege de putere a fost observată prin intermediul "proxy"-lor de avere în datele istorice referitoare la suprafețele de case din Egiptul antic [17] sau în numărul de iobagi aparținând nobililor din Ungaria medievală [18].

Prin venit ne referim la banii primiți pentru muncă, sau extrași din investiții [19], în timp ce prin avere ne referim la cantitatea de obiecte de valoare pe care o persoană o deține [20]. Atât valorile pentru venit cât și pentru avere sunt distribuite similar, și sunt acompaniate de legi de putere. Trebuie să remarcăm faptul că averea și venitul nu sunt însă atât de strâns legate pe cât cineva s-ar aștepta [21].

Analiza datelor

Studiile noastre experimentale ale veniturilor sunt bazate pe date exhaustive preluate din asigurările sociale din 2001 până în 2009 în județul Cluj, România [22]. De asemenea s-au preluat date fiscale din Ungaria, baze de date privind veniturile din sondaje publice pentru Statele Unite ale Americii [23] și pentru Finlanda [24], și statistici din anuarul statistic al Rusiei [25]. Totodată au fost colectate și folosite date fiscale publice pentru Australia [26] și Japonia [27].

Datele privind averea au fost obținute dintr-o singură sursă, din baze de date ale institutului World Inequality Database [28] privind averea estimată cu o metodologie complexă [29] din Statele Unite, Rusia și Franța.

Forma generală de distribuție a averii și a venitului

În ultimele decenii în econo-fizică, forma distribuției venitului și averii au fost studiate intens [30]. Densitatea de probabilitate (p.d.f.) este considerată în ambele cazuri ca o funcție cu caracteristici similare, pentru valori mari converge la o lege de putere, și prezintă un trend exponențial pentru valori medii. În studiile noastre considerăm o regiune adițională în p.d.f. în cazul valorilor mici sau negative. Vă prezentăm acestea pe Fig. 2.1. Cercetătorii din domeniul econo-fizicii au aplicat metodele fizicii statistice pentru o înțelegere mai profundă a acestei forme caracteristice. [30–33].

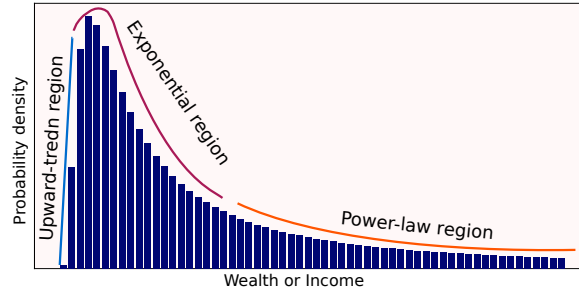


Figura 2.1: Représentarea schematică a celor trei regiuni definite în p.d.f.

Construirea unui model adecvat și realist

Presupunem în abordarea noastră că un model adecvat de distribuție a averii sau a venitului trebuie să se bazeze pe presupuneri realiste. Modelul trebuie să producă p.d.f. care converge către legi de putere în cazul valorilor mari și poate să descrie averi negative fiind continuu pe intervalul întreg.

2.1 Abordarea ecuației master, Creșterea Locală și Resetare Globală

Studiile noastre se bazează pe abordarea ecuației master [15]. Această ecuație evolutivă se bazează pe o creștere locală între stări consecutive și o resetare globală la o stare nouă. Abrevierea **LGGR**, vine din limba engleză (Local Growth and Global Reset). Ecuația evoluției pentru probabilități $P_k(t)$ (probabilitatea ca o componentă a sistemului să dețină k -cuante) poate fi scrisă prin ecuația:

$$\frac{dP_k(t)}{dt} = \delta_{k,0}\bar{\lambda}(t) - \lambda_k P_k(t) + v_{k-1}P_{k-1}(t) - v_k P_k(t) \quad (2.2)$$

Termenul v_k în Eq. 2.2. reprezintă rata de creștere de la starea cu k cuante deținute, la starea cu $k + 1$ cuante, în timp ce λ_k reprezintă rata de resetare către starea în care $k = 0$. Procesul este prezentat în Fig. 2.2. Generalizând ecuația evoluției pentru cazul stărilor

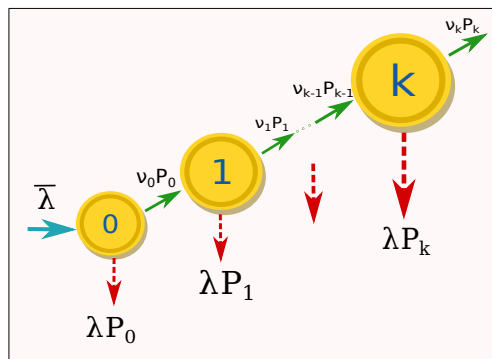


Figura 2.2: Vizualizarea procesului LGGR între stările descrise în Eq.2.2.

continue p.d.f.-ul staționar va fi [15, 36]:

$$\rho_s(y) = \rho_s(0) \frac{v(0)}{v(y)} e^{-\int_0^y \frac{\lambda(z)}{v(z)} dz} \quad (2.3)$$

2.1.1 Aplicarea metodei LGGR pentru descrierea funcției de distribuție a venitului

S-a utilizat baze de date exhaustive referitoare la județul Cluj, România, pentru studierea dinamicii veniturilor. Au fost urmărite veniturile individuale ale angajaților preluate din bazele de date pentru a înțelege dinamica sistemului. Prin observarea creșterii anuale a salariilor (Fig.2.3) s-a observat o funcție linear crescătoare pentru rata de creștere $v(y)$, (Eq.2.4).

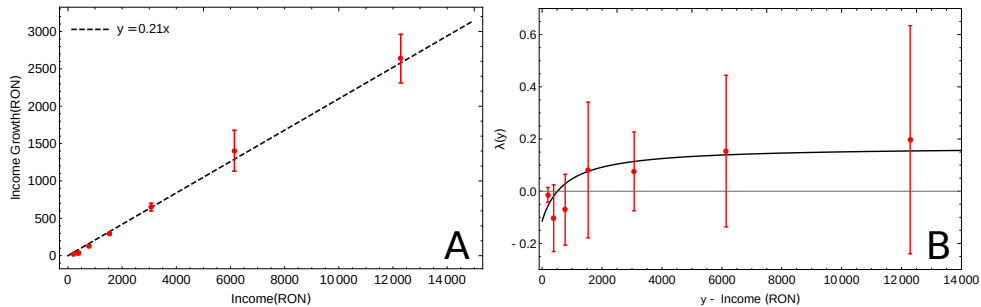


Figura 2.3: A: Rata de creștere în funcție de venitul mediu. Pe axa orizontală este prezentat venitul mediu în RON a k -a grupă a angajaților, pe axa verticală modificarea medie anuală a venitului pentru a k -a grupă ($\Delta_k i$). B: Ratele de resetare determinate experimental. Pe axa orizontală vă prezentăm venitul mediu în RON a k -a grupa angajaților, pe axa verticală rata de resetare medie $\lambda(y)$ pentru a k -a grupa. Linia neagră reprezintă regresia folosită. Abaterea standard de la valoarea medie a punctelor experimentale este reprezentat cu bare roșii.

$$v(y) = \beta(y + h) \quad (2.4)$$

Experimental s-a determinat rata de resetare $\lambda(y)$, urmărind numărul angajaților care intră și care se pensionează din forța de muncă în fiecare interval de venit (Fig.2.3). Rata aceasta poate fi aproximată prin Eq. 2.5 (rata de resetare "smart"), o funcție negativă pentru venituri mici, și pozitivă pentru valori mari.

$$\lambda(y) = S - \frac{a}{y + c} \quad (2.5)$$

Folosind ratele de creștere și resetare determinate experimental se calculează p.d.f.-ul pentru venituri scalate cu venitul mediu anual:

$$\bar{y}\rho_s(y) = \left(\frac{m-n}{n-1}\right)^{m-n} \frac{\Gamma(m)}{\Gamma(m-n)\Gamma(n)} \left(1 + \frac{y}{\bar{y}} \frac{m-n}{n-1}\right)^{-m} \left(\frac{y}{\bar{y}}\right)^{m-n-1} \quad (2.6)$$

Prin rescalarea veniturilor la venitul mediu al anului investigat, distribuțiile experimentale colapsează pe o singură curbă, care poate fi descrisă cu Eq. 2.6. Aceasta este o distribuție de tip Beta Prime. Pentru județul Cluj distribuția are următoarea formă

simplă:

$$\bar{y}\rho_s(y) = 12\frac{y}{\bar{y}}\left(1 + \frac{y}{\bar{y}}\right)^{-5} \quad (2.7)$$

Au fost construite distribuțiile experimentale pentru Ungaria și Japonia (Fig. 2.7). Cu ipoteza fenomenologică ca $n = m - 2$ distribuțiile acestea pot fi descrise teoretic cu un singur parametru m .

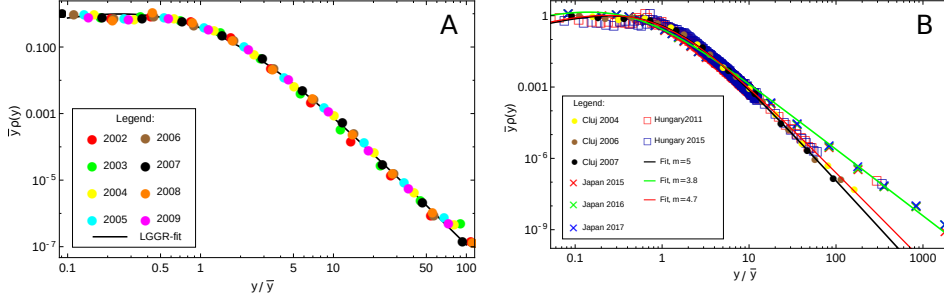


Figura 2.4: *A: Distribuția experimentală pentru venitul rescalat cu media anuală pentru județul Cluj. Curba neagră este p.d.f.-ul teoretic din Eq. 2.7. B: Distribuția experimentală pentru venitul rescalat cu media anuală pentru județul Cluj, Japonia și Ungaria. Pentru județul Cluj $m = 5$, pentru Japonia $m = 3.7$ și pentru Ungaria $m = 4.7$.*

2.1.2 Aplicarea metodei LGGR pentru descrierea funcției de distribuție a averii

Valorile averii au fost normalizate cu media anuală $r = \frac{R}{R}$ (în mod similar cu venitul). Distribuțiile rezultate colapsează pe aceeași curbă. Ca exemplu distribuția averii este prezentată pentru SUA pe Fig. 2.5.

Rata de creștere liniară pentru dinamica averii

Între anii 2001-2019 s-a studiat procesul de creștere a averii pentru 22 de persoane din lista de miliardari Forbes [34]. S-a observat o creștere preferențială liniară, similară cu cea din cazul veniturilor (Eq. 2.4). Presupunem în mod intuitiv că rata de resetare are aceeași formă ca cea utilizată în modelarea distribuției veniturilor (Eq.2.5).

Considerăm condiția $c = h$ în Eq. 2.4 și Eq. 2.5. Această condiție permite averii să crească pe interval $[-h, \infty)$, și resetarea să introducă indivizi cu avere negativă. Folosind ratele acestea în Eq. 2.3, s-a obținut p.d.f.-ul staționar (Eq. 2.8).

$$\rho_s(y) = \frac{a^S}{\Gamma(S)} e^{-\frac{a}{h+y}} (h+y)^{-1-S} \quad (2.8)$$

Normalizăm cu averea medie în Eq. 2.8, și folosim notația $r = \frac{y}{\bar{y}}$ pentru variabila de avere. Folosind notația $\kappa = \frac{a}{h(S-1)} - 1$ și $\bar{y} = \kappa h$, p.d.f.-ul averii normalizat cu averea medie se găsește sub forma Eq. 2.9.

$$\rho(r) = \frac{\kappa(\kappa+1)^S(S-1)^S}{\Gamma(S)} e^{-(\kappa+1)\frac{S-1}{1+\kappa r}} (1+\kappa r)^{-1-S} \quad (2.9)$$

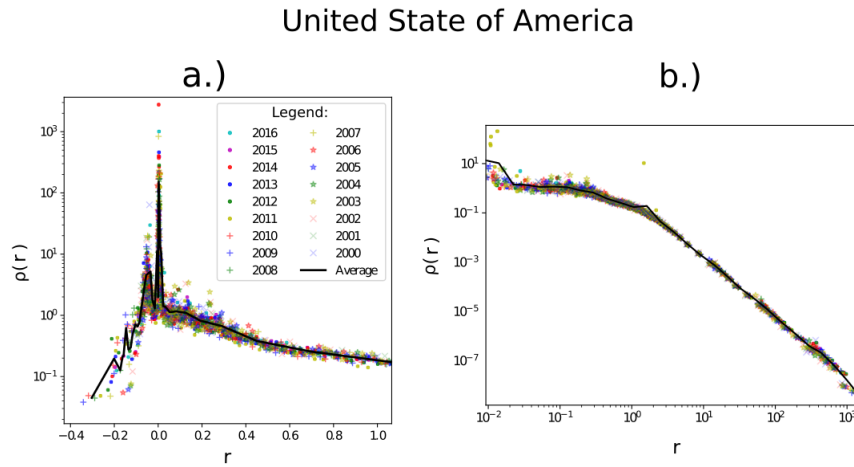


Figura 2.5: Distribuția experimentală pentru avere rescalată cu media anuală $r = \frac{R}{R}$ pentru SUA. Cu linia neagră se prezintă media p.d.f.-urilor anilor diferiți. **a.)** Capătul inferior al distribuției (sub valoarea medie) și regiunea cu avere negativă (datoria) sunt prezentate pe o scară log-normală. **b.)** Partea pozitivă a distribuției de avere este prezentată pe o scară dublu logaritmică.

Scalarea cu averea medie redefinieste domeniul pe care este definit p.d.f.-ul:

$$[-h, \infty) \rightarrow \left[-\frac{1}{\kappa}, \infty\right) \quad (2.10)$$

Comparație cu rezultatele experimentale

Folosind p.d.f.-ul stabilit anterior (Eq. 2.9) vom încerca să descriem distribuțiile observate experimental. Pentru Statele Unite și Rusia prezentăm rezultatele experimentale medii în Fig. 2.6, împreună cu p.d.f.-ul teoretic, cu parametrii $S = 1.4$ și $\kappa = 6.5$. Regresia cu acești parametri prezintă o descriere cantitativă bună, pentru fiecare regiune a averii.

Modelul nostru poate fi comparat cu modelul Mezard & Bouchaud [35] (MB) (care este o abordare bazată pe schimbul de avere între agenți și pe un proces de creștere aleatorie multiplicativă). Modelul MB are un singur parametru:

$$\rho_{MB}(r) = \frac{k^k}{\Gamma(k)} e^{-\frac{k}{r}} r^{-2-k} \quad (2.11)$$

Compararea modelului LGGR și MB cu datele experimentale sunt prezentate în Fig. 2.7. În cazul Franței, ambele modele produc aceleași legi de putere pentru distribuție, cu un exponent similar: -2.68 , deci ambele modele descriu bine regiunea pozitivă a averii. Abordarea MB este valabilă numai pe $[0, \infty)$. P.d.f.-ul obținut din modelul LGGR pentru avere în acest caz implică existența unei cantități neglijabile de avere negativă: $F_{neg} \propto 0.063$ pentru parametrii $\kappa = 7$ și $S = 1.68$, și descrie mai bine regiunea inferioară a distribuției averii.

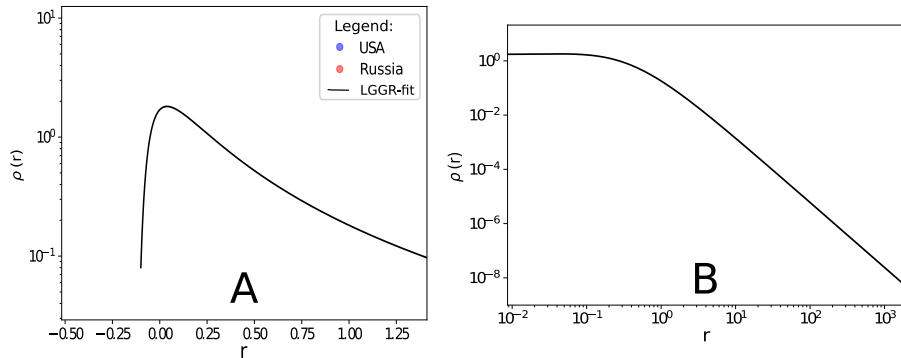


Figura 2.6: Funcția densității de probabilitate a averii normalizate cu media pentru SUA și Rusia. Simbolurile reprezintă p.d.f.-ul mediu pentru diferiți ani. Trenduri similare sunt observate pentru SUA și Rusia. Trendul aceasta poate fi descris cu Eq. 2.9 prin alegerea parametrilor $S = 1.4$ și $\kappa = 6.5$. **A:** Capătul inferior al distribuției (sub valoarea de medie) și regiunea cu avere negativă (datoria) sunt prezentate pe o scară log-normală. **B:** Partea pozitivă p.d.f.-ul de avere este prezentată pe o scară dublu logaritmic.

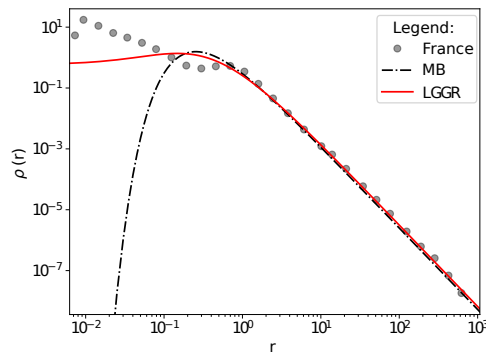


Figura 2.7: Aplicarea modelului LGGR și MB pentru cazul Franței. Parametrii fit-ului în abordarea LGGR sunt $\kappa = 7$, $S = 1.68$, iar pentru abordare MB parametrul este $k = 0.86$.

2.2 Concluzie și discuție

Prin examinarea datelor experimentale, și prin modelare teoretică s-a dezvăluit legi statistice interesante pentru distribuția venitului și a averii. Prin renormalizarea valorilor de venit și avere cu media anuală a acestora s-a observat o universalitate interesantă, colapsul p.d.f.-ului pe o curbă bine definită caracteristic țărilor studiate. Se observă o similaritate remarcabilă în distribuția averii pentru Statele Unite și Rusia, ceea ce ar putea fi însă doar o coincidență statistică.

Folosind date exhaustive s-au identificat legi statistice a dinamicii veniturilor. Aceste dinamici au fost descrise cu o rată de creștere lineară. Dinamica de resetare observată în sistem poate fi descrisă cu o rată de resetare inteligentă, care este bine susținut de date experimentale.

Folosind ratele anterior menționate în modelul LGGR s-a obținut o distribuție realistă a venitului. Aceleași rate cu modificări minore au fost folosite și pentru a descrie

distribuția averilor în societățile moderne.

Prin ipoteze simple și intuitive, modelul nostru generează funcții de distribuție calitativ similare, cu cele observate pentru venit și avere.

Metodele de modelare LGGR au însăși anumite deficiențe: nu iau în considerare inflația și creșterea economică, efecte care pot fi observate în datele experimentale.

Studiu de caz pentru estimarea averii și modelarea acesteia

Metodele utilizate pentru estimarea averii folosite de institutul WID [37] a căror date au fost folosite în studiul anterior au multe deficiențe, nefiind date exhaustive. Ca o consecință, rezultatele prezentate ar putea fi și ele influențate de acest lucru.

În literatura de specialitate lipsesc studiile referitoare la avere bazate pe date exhaustive. Există și o lipsă a studiilor de avere în sistemele economice diferite de cele din piața liberă. În capitolul acesta ne propunem un studiu în acest domeniu, de importanță primordială în econo-fizică.

Date exhaustive despre avere folosind registre agricole și evidențe fiscale

S-a creat un set de date exhaustive privind averea, care acoperă epoci cu diferite condiții economice. Autoritățile locale din comuna Sâncraiu (județul Cluj, România) au permis accesul la Registrele Agricole pentru anii 1961-1989. De asemenea ne-au permis accesul la datele fiscale pentru anul 2021 păstrând anonimatul persoanelor. Registrele Agricole au fost digitalizate de către echipa noastră.

Comuna Sâncraiu este formată din cinci sate, și se găsește la sud de orașul Huedin, România. În Fig. 3.1. este prezentată o hartă administrativă a comunei.

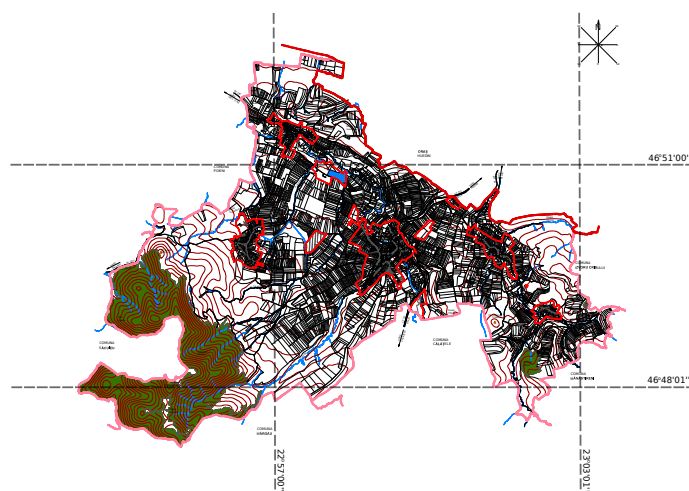


Figura 3.1: Harta administrativă a comunei Sâncraiu. (Sursa: Autorități locale.)

Din cauze geografice terenul este potrivit doar pentru agricultură limitată, iar din punct de vedere istoric, localnicii au dus un stil de viață preponderent agricol. Se poate presupune că într-o astfel de societate, averea constă în suprafața pământului deținut, cât și animalele, casele și grajdurile aflate în posesia gospodarilor.

3.1 Metodologia de estimare a averii

Bunurile agricole împărțite pe categorii, deținute de fiecare gospodărie a comunei sunt cuprinse în registrele agricole din anii 1961-1989. Exemple pentru aceste categorii sunt: dimensiunea terenului, a casei și a clădirilor auxiliare deținute, numărul animalelor după tip etc. Pentru a estima averea deținută de o singură gospodărie W . S-a calculat suma ponderată w_i a categoriilor de bunuri C_i din registrele agricole:

$$W = \sum C_i w_i \quad (3.1)$$

Valoarea relativă a diferitelor categorii de bunuri (una față de alta) este descrisă prin factorii de ponderare. O măsură unificată a averii (care poate fi considerată ca un fel de "monedă") este creată prin însumarea categoriilor de bunuri cu o pondere corespunzătoare. În continuare ne referim la această măsură unificată a averii ("moneda") ca "unități de avere" (wealth units $w.u.$). Categoriile de bunuri documentate în registrele agricole se schimbă de-a lungul anilor, care au ca și consecință directă schimbarea valorii lor exprimate în $w.u.$.

Datorită caracterului agricol al zonei, putem considera că registrele agricole documentează în mod realist averea gospodăriilor în anii 1969-1989. Date fiscale au fost folosite pentru estimarea averii individuale în anul 2021. Menționăm faptul că registrele agricole și datele fiscale nu conțin informații privind datoriile (averea negativă). Estimarea averii pentru anul 2021 a fost realizată cu metoda menționată anterior, considerând categorii diferite de bunuri (valoarea clădirilor, terenuri deținut).

Folosind un set de parametri de ponderare este posibilă supraestimarea valorii a anumitor categorii de bunuri. În metoda noastră de estimare mai multe seturi realiste de parametri de ponderare au fost luate în considerare. Au fost luate în analiză 10 seturi de parametri cu ponderi diferite, pentru fiecare an studiat. Aceasta a permis construirea barelor de eroare în distribuția averii estimate.

Proporția diferitelor bunuri din averea totală este puternic influențată de valoarea parametrilor de ponderare. Considerând compozițiile diferite ale averii totale la nivelul comunei, s-a putut reduce influența erorilor în estimarea averilor individuale. Efectul parametrilor diferiți este detaliat în tabelele dedicate din teză.

Averile individuale ale gospodăriilor au fost estimate cu metoda această pentru anii 1961, 1989 și 2021. Prin utilizarea metodei histogramelor, s-a estimat distribuția averii gospodăriilor $\rho(W)$. Efectul diferitelor seturi de ponderare se manifestă prin deplasarea mijlocului intervalelor (bins) din histogramă pe axa x și y . Din aceste deplasări se pot determina barele de eroare, care reprezintă abaterea de la valoarea medie. Prin această metodă este posibil să prezentăm efectul seturilor de ponderare în mod simultan. Rezultatele sunt prezentate în Fig. 3.2.

Această metodă de reprezentare arată o tendință clară cu o incertitudine relevantă. Forma distribuției nu este afectată semnificativ de diferitele seturi de ponderare, deși locația exactă a punctelor de date se schimbă. Vom reveni la această observație într-o secțiune ulterioară.

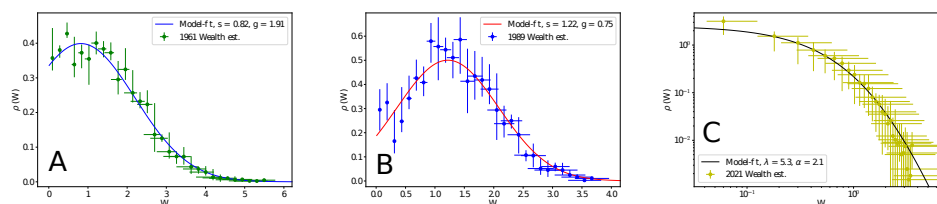


Figura 3.2: Funcția de densitate de probabilitate normalizată a averii pe anii 1961 (A), 1989 (B), și 2021 (C). Barele de pe puncte reprezintă variațiile obținute din diferite seturi de ponderare. Linia continuă reprezintă regresia obținută din metoda de modelare LGGR.

3.2 Abordare teoretică pentru descrierea distribuțiilor observate

Metoda de modelare LGGR oferă o descriere teoretică a dinamicii distribuției de avere în diferite condiții economice.

Averea privată sub regimul comunist? - Rată de creștere constantă cu rată de resetare liniară

Efectele socio-economice ale politicii comuniste pot fi înțelese prin rate adecvate de creștere și resetare din punctul de vedere al metodei LGGR. Creșterea averilor gospodăriilor este înceată, independent de averea inițială a gospodăriei. Aceasta poate fi descrisă cu o rată de creștere constantă $v(x) = g$. Putem considera o creștere controlată, pentru fiecare element al societății.

În sistemul comunist putem defini o cantitate ideală de avere privată, care este controlat de sistem. Descriem efectul aceasta cu o funcție lineară pentru rata de reset $\lambda(x) = x - s$. Cu aceste rate putem să calculăm p.d.f.-ul staționar al averii:

$$\rho_{s_{comm}}(x) = \sqrt{\frac{2}{g\pi}} \left[1 + \operatorname{erf}\left(\frac{s}{\sqrt{2g}}\right) \right]^{-1} e^{-\frac{x(x-2s)}{2g}} \quad (3.2)$$

Vă prezentăm regresia $\rho_{s_{comm}}$ împreună cu rezultatele experimentale în Fig.3.2. pentru anii 1961 și 1989. Parametrii de regresie, mediile experimentale și teoretice sunt prezentate în Tabelul 3.1. Valoarea averii medii este și valoarea parametrilor de reset - s , mai

Year.	g	s	R^2	\bar{x}_{theo}	\bar{x}_{exp}
1961	1.91	0.82	0.98	1.459 w.u.	[1.36; 1.49] w.u.
1989	0.75	1.22	0.93	1.359 w.u.	[1.199; 1.391] w.u.

Tabel 3.1: Parametrii de regresie în Eq.3.2 pentru anii 1961, 1989. Mediile calculate experimental \bar{x}_{exp} și teoretic \bar{x}_{theo} .

ales în cazul anului 1989.

Dinamica cu creștere preferențială - Rată de creștere liniară cu rată constantă de resetare.

Pentru modelarea dinamicii averii în anul 2021, s-a reconsiderat abordarea utilizată în Capitolul 2 pentru economia de piață liberă regăsită în societățile moderne. În acest caz, însă, nu există persoane cu datorii (avere negativă) în comună. Folosind rata de creștere liniară $v(x) = x + \alpha$ pentru a descrie caracterul preferențial al interacțiunilor economice, s-a considerat o rată de resetare constantă $\lambda(x) = \lambda$. Aceasta înseamnă că resetarea are aceeași probabilitate pentru fiecare individ, independent de averea deținută. În limita staționară p.d.f-ul are forma următoarea [15]:

$$\rho_{scap} = \frac{\lambda}{\alpha} \left(1 + \frac{x}{\alpha}\right)^{-1-\lambda} \quad (3.3)$$

Această este o distribuție Tsallis-Pareto (sau Lomax tip II.) care este o distribuție așteptată în sistemele cu dinamică de creștere preferențială. Vă prezentăm regresia ρ_{scap} împreună cu rezultatele experimentale în Fig.3.2. Parametrii de regresie, mediile experimentale și cele teoretice sunt prezentate în Tabelul. 3.2.

Year.	λ	α	R^2	\bar{x}_{theo}	\bar{x}_{exp}
2021	5.3	2.1	0.91	0.488 w.u.	[0.297;0.682] w.u.

Tabel 3.2: Parametrii aleși pentru regresia Tsallis-Pareto din Eq. 3.3. pentru anul 2021. Mediile calculate experimental \bar{x}_{exp} și teoretic \bar{x}_{theo} .

Mărimea eșantionului: în 1961 un total de 1133 de gospodării individuale, în 1989 aproximativ 921 de gospodării și în 2021 un total de 1595 de persoane.

3.3 Inegalități observate

S-a calculat valoarea celor mai cunoscute măsuri de inegalitate bazate pe rezultatele experimentale și teoretice. În Tabelul 3.3. prezentăm coeficientul de Gini G [38], Punctul-Pareto P_p (proporția P_p din societate care deține $1 - P_p$ din proporția totală a averii) comparând valorile experimentale și cele oferite de modelul LGGR. Instrumentul de vizualizare a inegalității - curba Lorenz (proporția cumulativă a averii $F(x)$ în funcție de proporția cumulativă a societății care o deține $C(x)$) [39] este prezentat în Fig. 3.3.

Year	G		P_p	
	Theoretical	Experimental	Theoretical	Experimental
1961	0.378	[0.377;0.379]	0.367	[0.366;0.368]
1989	0.312	[0.304;0.315]	0.391	[0.390;0.395]
2021	0.552	[0.543;0.579]	0.298	[0.282;0.299]

Tabel 3.3: Coeficientul Gini și Punctul-Pareto pentru anii 1961, 1989, 2021. Valorile experimentale sunt calculate din estimările de avere, în timp ce valorile teoretice sunt calculate din p.d.f-urile obținute prin modelul LGGR.

Valorile inegalității arată o scădere între 1961 și 1989. După cum este de așteptat, economia de piață liberă a dus la o creștere a inegalității până în anul 2021. Aceste schimbări sunt reflectate prin p.d.f.-urile rezultate din modelul LGGR.

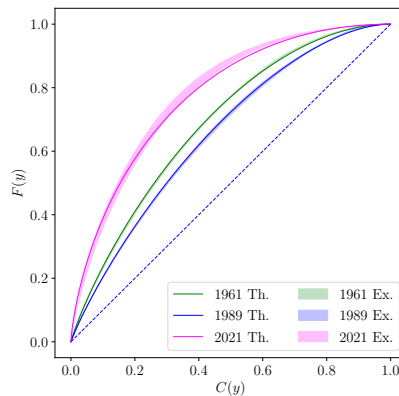


Figura 3.3: Curbe Lorenz pentru anii 1961, 1989, 2021. Rezultatele experimentale sunt reprezentate cu o regiune umbră. Linia continuă reprezintă estimarea teoretică folosind rezultatele modelului LGGR.

3.4 Concluzie și discuție

Distribuțiile averii observate sunt reprezentative pentru trei perioade socio-economice diferite ale României. În anul 1961, guvernul comunist a impus colectivizarea iar structurile de avere existente în acest an sunt rezultatul reformelor agrare din anii 1921 și 1945. Anul 1989 este ultimul an al regimului comunist din România, colectivizarea reducând averea privată în mare măsură. Ultimul an studiat este 2021, la 32 de ani de la sfârșitul comunismului, după trecerea la o economie capitalistă. Funcția de densitate a probabilității de avere pentru anii 1961 și 1989 este diferită de cea pe care s-a observat în societățile moderne. Pentru acești ani distribuția de avere are o formă asemănătoare cu cea Gaussiană. Pentru anul 2021, distribuția averii este o lege de putere caracteristică, similară cu observațiile lui Pareto.

Utilizarea seturilor de ponderare a fost crucială pentru estimarea corectă a averilor, neavând informații cu privire la valorile de piață exacte ale categoriilor de bunuri. Această metodă de suprapunere a estimărilor multiple ne-a permis să tragem concluzii calitative cu privire la forma distribuției averii, și în mod remarcabil, nu a modificat tendințele calitative ale distribuției. Această stabilitate a formei observate pentru funcția de distribuție experimentală ar putea fi consecința faptului că economia din 1961 și parțial cea din 1989 s-a bazat pe agricultură. Presupunem că există o relație directă între categoriile de bunuri, deoarece suprafața terenului deținută determină cantitatea altor bunuri agricole deținute (animale, etc.). Comerțul liber cu terenuri era inexistent în anii dinainte de 1989. Cea mai mare parte a veniturilor gospodăriilor era investită în obiectul cel mai valoros și încă privat: casa. Acest lucru se observă în date, suprafața totală a caselor private în 1961 fiind de 44273 m^2 și care a crescut la 65396 m^2 în 1989. În anul 2021 economia locală nu mai este dominată de agricultură deoarece au apărut noi surse de venit care creează avere, precum turismul și industria de servicii. Valorile medii experimentale ale averii sunt mai răspândite, lucru care rezultă în lărgirea barelor de eroare din grafic. Acesta este probabil rezultatul numărului redus de categorii din date fiscale și al diversificării economiei locale. O explicația simplă este următoarea: categoriile de bunuri au fost decuplate de valoarea terenului.

Aproximarea LGGR și justificarea ratelor

Considerând ratele de creștere și resetare realiste din modelul LGGR, s-a reprodus p.d.f.-ul averii atât pentru anul 1961, cât și pentru anul 1989. Dinamica realizată de procesele considerate reflectă într-un mod simplificat situația economică din regimul comunist. Observăm scăderea inegalităților din anul 1961 până în anul 1981, cât și scăderea parametrilor g odată cu creșterea ratei de resetare s (Tabelul 3.1). Funcția de densitate atinge o valoare maximă pentru valori de avere egale cu parametrul s (valoarea averii dorite). Odată cu creșterea valorii parametrului s , maximul distribuției se apropie de valoarea medie a averii, reducând inegalitatea din sistem.

Presupunem că în anul 2021, dinamica acumulării de avere este creșterea preferențială (principiul Matei), având o rată de creștere lineară (descrișă în detaliu în capitolul anterior). Faptul că niciunul dintre locuitorii satului nu avea datorii, ne-a permis să alegem o rată constantă simplă de resetare. Această abordare rezultă într-o distribuție de tip Tsallis-Pareto. Exponentul „ a ” din legea de putere în distribuția averii obținute pentru comuna Sâncraiu în anul 2021 poate fi comparat cu cel obținut pentru societățile moderne la nivel de țară (lucru explicat în capitol trecut). Această comparație este prezentată în Tabelul 3.4.

Wealth Data source	Power-law exponent a
Wealth data France (2000-2014)	2.68
Wealth data USA (2000-2016) and Russia (2000-2015)	2.4
Sâncraiu commune Tax data (2021)	6.3

Tabel 3.4: Exponentul „ a ” din legea de putere în distribuția averii pentru SUA, Franța și Rusia (extras din Capitolul. 2), comparativ cu exponentul „ a ” din distribuția averii pentru comuna Sâncraiu în anul 2021.

În cazul aceasta, exponentul este mai mare (ceea ce înseamnă că inegalitățile sunt mai mici) pentru comuna Sâncraiu decât cele observate la nivel de țară. Diferența poate fi explicată prin numărul relativ mic de locuitori al comunei și lipsa categoriei persoanelor foarte bogate.

În acest capitol s-a prezentat o abordare „bottom-up” pentru estimarea averii folosind „proxy”-uri simple. Folosind rate de creștere și resetare pentru modelul LGGR, justificate din punct de vedere economic, s-a descris distribuția în diferite situații economice. Rezultatele teoretice reflectă observațiile experimentale cantitative, cum ar fi forma distribuției, averea medie și inegalitatea.

Subliniem faptul că autorii acestei lucrări nu se aliniază cu politicile sau ideologiile politice care ar avea legătură cu această lucrare. Ne limităm pur și simplu să prezentăm faptele așa cum au fost observate din evidențele agricole și fiscale de-a lungul timpului.

Modelarea topologiei fizice de interconectare în rețeaua de Internet

Internetul se bazează pe o rețea fizică de computere interconectate. S-a demonstrat că numărul de conexiuni la Internet pe care le are un oraș este direct legat de factori socio-economici precum PIB-ul [40].

În acest capitol prezentăm o lege de scalare dinamică și un model simplu de rețea care poate să genereze grafuri bazate pe considerații realiste, care au proprietăți similare cu cele ale rețelei de Internet.

Despre Internet

Internetul este o infrastructură fizică care interconectează rețelele de calculatoare [41]. În cadrul modelului OSI [42], considerăm că la nivelul de bază Internetul funcționează ca o rețea de cabluri între computere și routere (între care se trimit pachete de date).

4.1 Proprietăți dinamice și topologice relevante ale Internetului

Internetul poate fi considerat un sistem complex care prezintă câteva proprietăți de scalare interesante. S-a descoperit recent o nouă lege de putere - scalarea [16] între timpul dus-întors (Round-Trip-Time sau RTT) al unei solicitări Ping și distanța geografică dintre două computere [43, 44] legate la rețea. Solicitarea Ping este utilizată la măsurarea timpului de acces la un dispozitiv aflat la o anumită distanță în rețeaua Internet. Un pachet de date este transmis către un dispozitiv aflat la o anumită distanță. Acest dispozitiv va trimite înapoi un răspuns de confirmare. Calculatorul expeditor măsoară timpul scurs în milisecunde de la cerere până la răspuns. În experimentul considerat, 24700 de dispozitive având coordonate distribuite pe tot globul, au fost abordate prin „Ping” pentru a obține date. Rezultatele referitoare la RTT au fost mediate.

Pentru a avea consecvență în date, acest studiu a fost replicat cu ajutorul instrumentului **Traceroute**, care este similar cu „Ping”, dar include și alte funcții suplimentare. Solicitarea dată de către Traceroute urmărește traseul pachetelor de date, de la punctul de plecare către destinație, returnând un pachet de notificare de la fiecare router care a fost traversat de date pe parcurs [45]. Un RTT poate fi calculat pentru fiecare pas al traseului. A fost utilizat setul de date IPv4 Routed/24 Topology Dataset furnizat de CAIDA [46] pentru anul 2007. Această bază de date a permis crearea unui graf dintr-o

secțiune redusă a Internetului, cu un număr de $V = 43992$ vârfuri și $C = 191005$ legături (conexiuni) între routere, cu un număr mediu de conexiuni pe noduri $\bar{c} = 8.68$. RTT-urile au fost extrase din acest set de date.

Experimentul Traceroute a permis studiarea scalării RTT-ului în raport cu numărul de routere intermediare ("pași"- "steps"). Scalarea RTT-ului în funcție de distanța fizică (l) și scalarea în funcție de numărul de „pași” ("steps") pe routere intermediare (S) sunt prezentate în Fig. 4.1.

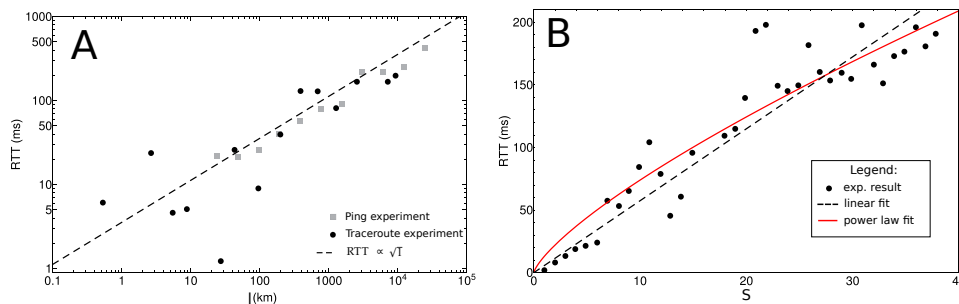


Figura 4.1: (A): RTT-ul mediu (în ms) este reprezentat în funcție de distanța fizică l (km). Sunt prezentate rezultatele metodelor Ping și Traceroute. Se observă proporționalitatea: $RTT \propto \sqrt{l}$. (B): RTT-ul mediu (în ms) în funcție de numărul de „pași” ("steps") al routerelor intermediare S obținut din datele Traceroute. S-a considerat două tipuri de regresii, unul liniar (linia întreruptă) și unul de o funcție de putere cu un exponent $\frac{3}{4}$. Regresia funcției de putere are coeficientul $R^2 = 0.98$.

Rezultatele experimentelor Traceroute și Ping pot fi combinate. Aceste legi de scalare sunt prezentate în Fig. 4.1. Observațiile pot fi rezumate astfel:

$$\begin{aligned} RTT &\propto S^{\frac{3}{4}} \\ RTT &\propto l^{\frac{1}{2}} \end{aligned} \quad (4.1)$$

Din Eq. 4.1. putem să obținem scalarea numărului de routere intermediare (S) în funcție de distanța geografică (l):

$$S \propto l^{\frac{2}{3}} \quad (4.2)$$

4.2 Modelarea topologiei de rețea a Internetului

Pentru a înțelege observațiile prezentate anterior propunem un model simplu care generează grafuri similare cu cele ale rețelei de Internet. În modelul nostru de rețea nodurile grafului sunt considerate a fi routerele atașate la orașe în timp ce conexiunile dintre noduri reprezintă conexiunile prin cablu dintre orașe. S-a observat în regiunea geografică studiată o proporționalitate directă între PIB-ul și populația orașelor. Din acest motiv, în model considerăm că numărul conexiunilor care aparțin la un oraș este proporțional cu populația acestuia. În cadrul modelului, orașele sunt considerate ca și puncte într-un spațiu simplu Euclidian cu coordonate aleatorii având o pondere realistă P_i reprezentând populația orașelor. Populația P_i urmează o distribuție de tip Tsallis-Pareto cu un exponent de scalare $a = -2$. Definim o zonă circulară imaginară de acoperire (extinderea maximă a posibilelor conexiuni) pentru fiecare oraș, având o rază proporțională cu rădăcina pătrată a populației: $R_i = \sigma\sqrt{P_i}$. În construcția rețelei

se consideră fiecare pereche posibilă de noduri și calculăm un factor de conectivitate $k_{ij} = \frac{R_i + R_j}{l_{ij}}$, unde l_{ij} este distanță geografică dintre nodurile i și j . În cazul $k_{ij} > 1$, o conexiune este creată între nodurile i și j . Acest proces de construire a grafului este prezentat în mod schematic în Fig. 4.2. În studiul nostru s-a luat în considerare doar cea mai mare componentă conectată a grafului generate (*giant component*), și s-au studiat proprietățile acesteia.

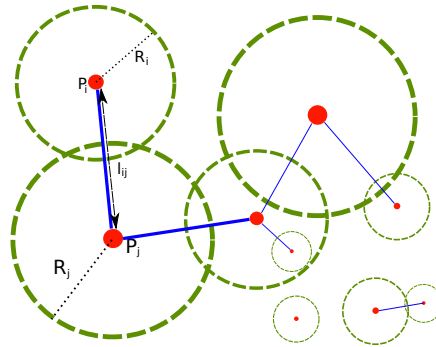


Figura 4.2: Reprezentarea schematică a modelului. Punctele roșii reprezintă nodurile, cercurile verzi sunt zona lor de acoperire, iar liniile albastre sunt conexiunile stabilite.

Proprietățile grafului generate de model

Experimentele computaționale au arătat că pentru parametrul $\sigma = 0.4$ numărul mediu de conexiuni pe noduri al întregului sistem generat a fost $\bar{c} \sim 8$. Această valoare este foarte apropiată de cea observată în experimente. S-a studiat proprietățile medii ale grafului având numărul de noduri fixe $M = 2400, 8000$ și parametrul $\sigma = 0.4$ cu o ansamblu de 100 de experimente computaționale. Proprietățile grafului generate de modelul nostru și cele observate din datele Traceroute au fost similare cu valorile regăsite în literatură [47]. Acest lucru este prezentat în Tabelul. 4.1.

Methods	Properties				
	Numărul Nodurilor (V)	Numărul Conexiunilor (C)	Numărul mediu de conexiuni pe vârf \bar{c}	Numărul maxim de conexiuni pe vârf C_m	Exponentul distribuției conexiunilor pe vârf γ $f(c) \propto c^{-1-\gamma}$
SKITTER [47]	9204	28959	6.29	2070	1.25
BGP [47]	17446	40805	4.68	2498	1.16
WHOOIS [47]	7485	56949	15.22	1079	-
Traceroute	43992	191005	8.68	10474	1.23
model, M=2400	1946	12607	12.95	1382	1.52
model, M=8000	6679	49748	14.88	4254	1.60

Tabel 4.1: Proprietățile grafului de Internet, cu diferite metode de cartografiere, în literatura de specialitate [47]. Acestea sunt comparate cu proprietățile observate în experimentul Traceroute și cu proprietățile medii obținute de modelul nostru.

Au fost comparate distribuția conectivității nodurilor $p(c)$ pentru grafului generate (mediate din ansamblu) și distribuția conectivității nodurilor pentru o subsecțiune

a Internetului (obținute din datele Traceroute). Internetul este considerat a fi o rețea fără scală [47, 48], deoarece distribuția caracteristică a conectivității nodurilor fiind o lege de putere. Pentru această distribuție o regresie bună este oferită de cea de tip Tsallis-Pareto (sau Lomax 2.) având media de conexiuni pe noduri $\bar{c} = 8.68$ obținută din experimentul Traceroute:

$$p(c) = \frac{\gamma}{\bar{c}(\gamma - 1)} \left(1 + \frac{c}{\bar{c}(\gamma - 1)}\right)^{-\gamma-1} \quad (4.3)$$

Distribuția conectivității nodurilor obținut din experimentul Traceroute și distribuția pentru grafuri creată de modelul nostru sunt prezentate în Fig. 4.3. Rezultatele modelului (pentru diferite dimensiuni) și rezultatele obținute din experimentul Traceroute coincid acceptabil. Grafurile generate cu modelul nostru au proprietăți topologice similare cu cele ale rețelei de Internet.

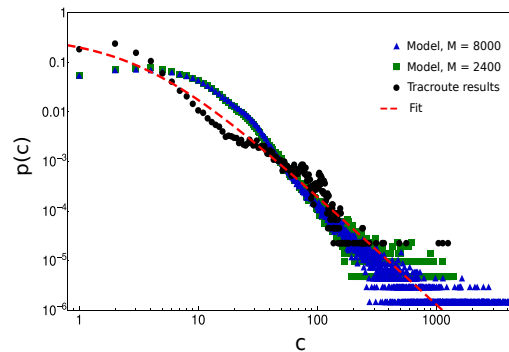


Figura 4.3: Distribuția conectivității nodurilor obținută din experimentul Traceroute și distribuția creată de modelul nostru având $M = 2400$ și $M = 8000$ noduri. Linia roșie întreruptă reprezintă distribuția de tip Tsallis-Pareto din Eq. 4.3 cu parametrii $\bar{c} = 8.68$ și $\gamma = 1.23$. Sursa: [16]

4.3 Scalare în procesul de propagare a informațiilor

Legea de scalare din Eq. 4.2. poate fi verificată folosind atât experimentul Traceroute, cât și grafurile generate de modelul nostru. Numărul de routere intermediare vizitate ("steps" sau "pași") în funcție de distanță l este afișat în Fig. 4.4. Pentru a compara rezultatele obținute din model cu rezultatele experimentului Traceroute, distanțele fizice experimentale au fost redimensionate la un pătrat cu dimensiuni 1×1 .

Cele mai scurte trasee pe graf pot fi calculate din grafurile generate de către modele. Acest lucru este analog cu minimizarea pașilor intermediari (prin politicile de rutare) în timpul transferului de date pe Internet. Prin similaritatea acestor procese, compararea rezultatelor experimentale cu rezultatele modelului este fezabilă.

4.4 Concluzie

În acest capitol s-a prezentat un model de rețea pentru descrierea topologică a conectivității Internetului (în spațiul fizic). Folosind modelul nostru s-a generat grafuri cu proprietăți topologice similare cu cele ale Internetului la nivel de router.

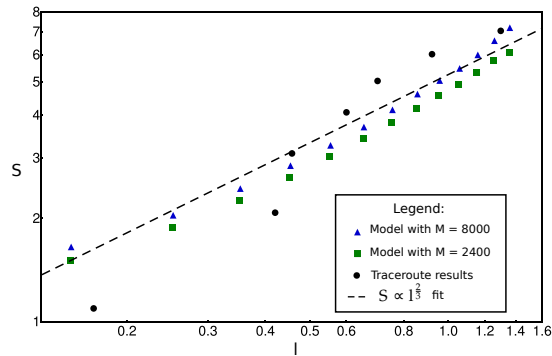


Figura 4.4: Numărul mediu de „pași” S reprezentat ca o funcție de distanță fizică l dintre sursă și destinație, pe o scară log-log. Rezultatele prezentate obținute din experimentul Traceroute au fost redimensionate la un pătrat cu dimensiuni 1×1 , pentru a putea fi comparate cu rezultatele din model. Rezultatele acestea arată scalarea $S \propto l^{\frac{2}{3}}$. Sursa: [16]

De asemenea, s-a prezentat legi de scalare cum ar fi durata medie de transport a informațiilor în raport cu distanța geografică. Modelul nostru a reprodus cu succes această scalare complexă $S \propto l^{\frac{2}{3}}$, care este regăsită și prin experimente.

Modelul nostru are niște deficiențe. Problema principală este că nu ține cont de o geografie realistă cum ar fi: limitele fizice sau politice care ar putea avea un efect asupra topologiei rețelei. Această abordare nu ia în considerare conexiunile ponderate, ceea ce înseamnă că bandwidth-urile conexiunilor reale de Internet nu sunt capturate deloc. Ca și consecință, în rețeaua modelată nu putem identifica o "coloană vertebrală a Internetului" care este o sub-rețea centrală cu bandwidth mare, ce asigură conexiuni de mare viteză pe distanțe lungi având întârziere redusă [49, 50].

În sfârșit putem concludiona că această abordare simplă descrie proprietățile calitative ale Internetului și poate fi considerată un prim-pas către o mai bună înțelegere a Internetului ca rețea.

Concluzie

"PSIHOISTORIA — ...Gaal Dornick folosind concepte non-matematice, a definit psihoistoria ca fiind ramura matematicii care se ocupă cu reacțiile conglomeratelor umane față de stimulii sociali și economici dați... În toate aceste definiții, este implicată presupunerea că conglomeratul uman în discuție trebuie să fie suficient de mare pentru a permite o tratare statistică valabilă."
- Isaac Asimov [51]

Citatul de mai sus este definiția „psihoistoriei”, un domeniu științific fictiv, piesa centrală a celebrei trilogii "Fundatia" a lui Asimov. Într-un fel acest citat încapsulează conținutul lucrării prezente. Putem declara că indivizii par să acționeze în moduri complet aleatorii, dictate de o logică internă misterioasă. Numărul mare de indivizi implică un număr și mai mare de opțiuni și interacțiuni posibile, prin urmare se creează un sistem complex de persoane pe care numim societate.

În trecut o înțelegere atât de profundă a societății era considerată "science-fiction". Lumea noastră devine din ce în ce mai digitalizată și odată cu aceasta, mai multe aspecte ale vieții noastre devin înregistrate. Această cantitate de date deschide posibilități nelimitate pentru a înțelege complexitatea societății noastre. Aceste descoperiri ne stârnesc curiozitatea, de a crea modele și de a explica aceste observații. În esența gândirii noastre, indivizii sunt analogi cu particulele gazelor, în timp ce societatea noastră este analogă cu gazul la nivel microscopic pentru care se aplică statisticile. Noi descriem distribuțiile observate, referitoare la dinamica societății noastre în mod similar cu distribuția Boltzmann care descrie energia particulelor din gaz.

Pornind de la modul de gândire descris mai sus, prima dată ne-am uitat la distribuția veniturilor pentru județul Cluj (România). Nu este surprinzător că aceasta este o distribuție care urmează o lege de putere. Caracterul exhaustiv al bazei de date ne-a permis să înțelegem dinamica internă a veniturilor din forța de muncă. Am observat o creștere preferențială a venitului, lucru care poate fi formulat ca "principiul Matei". Am studiat dinamica forței de muncă și din aceasta am reușit să tragem o altă concluzie: în medie, angajații încep cariera cu venituri mici și se pensionează cu venituri mai mari. Aceste constatări au fost formulate într-un mod matematic, ca rate de creștere și de resetare, în cadrul unui model evolutiv simplu, numit Creștere Locală și Resetare Globală (LGGR). Folosind această abordare am descris teoretic distribuția staționară a densității venitului. S-a dovedit că forma funcției de distribuție poate fi aplicată și pentru date de venituri din alte țări (cu alți parametri).

Din observațiile la nivelul indivizilor am ajuns la concluzii la nivelul sistemului (care sunt în concordanță cu observațiile empirice privind întregul sistem) reflectând modul de gândire descris în primele paragrafe ale concluziei. Am observat o universalitate interesantă: prin rescalarea venitului cu valoarea lui medie (pentru o anumită regiune sau țară) distribuțiile pentru diferiți ani colapsează.

Am studiat distribuția averii în societățile moderne cu date deja prelucrate de instituția WID. Am observat efectul menționat anterior (în cazul venitului), distribuția averii rescalate cu valoarea medie se regăsește pe aceeași curbă (pentru o țară dată).

Folosind abordarea LGGR am descris distribuția averii, folosind rate de creștere și resetare similare cu cele utilizate în cazul veniturilor (cu o ușoară modificare pentru a permite existența averii negative). Rezultatele noastre sunt comparabile cu alte abordări descrise în literatură.

În următorul studiu prezentat în această teză am propus o metodă simplă de estimare a averii bazată pe date exhaustive pentru comuna Sâncraiu. Metoda aceasta necesită cunoașterea, bunurile majore și relevante din punct de vedere socio-economic deținute de fiecare gospodărie (dintr-o anumită zonă geografică restrânsă). Dacă bunurile sunt clasificate în categorii predefinite, atunci putem pondera numărul bunurilor transformându-le valoarea într-o monedă ipotetică. De aici putem calcula combinația lor liniară pentru a obține o cantitate relevantă și unificată pentru descrierea averii unei gospodării. Pentru seturi de ponderi diferite, (care practic înseamnă compoziții de avere diferite) metoda noastră de estimare produce distribuții stabile ale averii, care sugerează o dependență între categoriile de bunuri. Această dependență a fost mai importantă în societatea agricolă și a devenit mai puțin semnificativă în economia modernă diversificată în aceeași comună (în ere economice diferite). Conform cunoștințelor noastre, această abordare de estimare a averii este nouă.

În acest studiu am descoperit în perioada comunistă o funcție de distribuție asemănătoare cu cea Gaussiană pentru avere. După 32 de ani de la abolirea comunismului, a apărut binecunoscuta distribuție Tsallis-Pareto pentru avere în aceeași comună (distribuție caracteristică pentru economia de piață liberă). Observațiile acestea au fost modelate folosind metoda LGGR, folosind rate justificate de politica economică a perioadelor comunistă și capitalistă.

Au fost calculate și discutate măsuri ale inegalității obținute din datele estimate de avere. Evoluția acestor măsuri de inegalitate au fost urmărite pentru comuna Sâncraiu în perioade cu condiții socio-economice diferite.

În capitolul final ne-am ocupat de un subiect oarecum diferit, dar acesta a fost urmărit printr-o metodologie similară de cercetare și modelare. Am studiat Internetul ca o rețea în spațiul fizic. Acest studiu este conectat de tema noastră principală deoarece creșterea Internetului ca rețea este influențată de factori socio-economici, cum ar fi PIB-ul. Folosind doar ipoteze bazate pe distribuția mărimii orașelor, am creat un model de rețea care este capabil să genereze grafuri cu proprietăți similare cu cele ale rețelei de Internet. Modelul nostru a poate să reproducă o scalarea observată între timpul de răspândire a informațiilor și distanța geografică.

Această lucrare și-a îndeplinit scopul inițial, deoarece a identificat distribuțiile caracteristice în sistemele socio-economice. S-au studiat observațiile experimentale din perspectiva sistemelor complexe și s-au aplicat metode de modelare pentru a explica teoretic aceste observații. Metodele prezentate în această teză sunt doar o primă încercare de a modela sistemele socio-economice folosind metodologia fizicii statistice. Pentru a înțelege legile de scalare observate, anumite aspecte ale acestor sisteme au fost ignorate.

Pentru viitor considerăm important să înțelegem conexiunile subtile dintre parametrii modelului nostru și parametrii socio-economici relevanți. În acest fel, ar fi posibilă promovarea lucrării noastre în comunitatea științelor sociale, atrăgând interes pentru metodologiile de modelare dezvoltate în econo-fizică.

Lista de publicatii

Publicații relevante pentru teză

- István Pap, Levente Varga, Mounir Afifi, **István Gere** and Zoltán Néda. *Scaling in the space-time of the Internet*. Scientific Reports, Vol. 9, 9734, (2019), IF = 4.380
- Zoltán Néda, **István Gere**, Tamás S. Biró, Géza Tóth and Noémi Derzsy. *Scaling in income inequalities and its dynamical origin*. Physica A: Statistical Mechanics and its Applications, Vol. 549, 124491, (2020), IF = 3.263
- **István Gere**, Szabolcs Kelemen, Géza Tóth, Tamás S. Biró, and Zoltán Néda. *Wealth distribution in modern societies: Collected data and a master equation approach*. Physica A: Statistical Mechanics and its Applications, Vol. 581, 126194, (2021), IF = 3.263
- **István Gere**, Szabolcs Kelemen, Tamás S. Biró and Zoltán Néda. *Wealth Distribution in Villages. Transition From Socialism to Capitalism in View of Exhaustive Wealth Data and a Master Equation Approach*. Frontiers in Physics, Vol. 10, 827143, (2022), IF = 3.560

Prezentări la conferințe

- István Pap, Mounir Afifi, Levente Varga, **István Gere** and Zoltán Néda. (14-17 June 2018) *Nontrivial dynamical scaling in the Internet: experiments and a simple model*. MaCS: 12th Joint Conference on Mathematics and Computer Science, Cluj-Napoca
- István Pap, Mounir Afifi, Levente Varga, **István Gere** and Zoltán Néda. (1-4 May, 2018) *Scaling in the space-time of the Internet*. MECO43: 43th Conference of the Middle European Cooperation in Statistical Physics, Krakow
- **István Gere**, András Kuki and Zoltán Néda. (14-16 September, 2020) *Towards probabilistic forecasting of earthquakes* [Online Conference, Poster presentation]. MECO45: 45th Conference of the Middle European Cooperation in Statistical Physics
- Zoltán Néda, **István Gere** and Szabolcs Kelemen (01-03 July, 2021) *The Growth and Reset model for social inequalities* [Online Conference, Oral presentation]. FENS 2021: 11th Polish Symposium on Physics in Economy and Social Sciences

Bibliografie selectată

- [1] Schinckus C. *When Physics Became Undisciplined: An Essay on Econophysics (Doctoral thesis)*. (Girton College, University of Cambridge, 2018)
- [2] Quetelet A. *Lettres à S. A. R. Le Duc Régnant de Saxe Cobourg et Gotha, sur la théorie des probabilités, appliquée aux sciences morales et politique*. (Brussels, Hayez, 1846)
- [3] Rice J. A. *Mathematical Statistics and Data Analysis*. (Belmont, CA: Duxbury Press, 2006)
- [4] Schmidt K.H. *A new test for random events of an exponential distribution*. *European Physical Journal A*, Vol. 8, 141–145, (2000)
- [5] Néda Z., Varga L., Biró T.S. *Science and Facebook: The same popularity law!* *PLoS ONE*, Vol. 12, e0179656. (2017)
- [6] Gutenberg R., Richter C.F. *Frequency of earthquakes in California*. *Bulletin of the Seismological Society of America*, Vol. 34, 185-188, (1944)
- [7] Rosen K. T., Resnick M. *The size distribution of cities: An examination of the Pareto law and primacy*. *Journal of Urban Economics*, Vol. 8, 165-186, (1980)
- [8] Varga L., Kovács A., Tóth G., Papp I., Néda Z. *Further We Travel the Faster We Go*. *PLoS ONE*, Vol. 11, e0148913. (2016)
- [9] Pareto V. *Cours D'Économie Politique*. (Macmillan, Paris, Vol. 2, 1897)
- [10] Kleinert H., Schulte-Frohlinde V. *Critical Properties of Φ^4 Theories*. (World Scientific Publishing Company, Singapore, 2001.)
- [11] Bak P., Tang C., Wiesenfeld K. *Self-organized criticality*. *Physical Review A*, Vol. 38, 364, (1988)
- [12] Newman M. *Power laws, Pareto distributions and Zipf's law*. *Contemporary Physics*, Vol. 46, 323-351, (2005)
- [13] Gere I., Kelemen S., Tóth G., Biró T. S., Néda Z. *Wealth distribution in modern societies: Collected data and a master equation approach* *Physica A*, Vol. 581, 126194, (2021)
- [14] Néda Z., Gere I., Biró T. S., Tóth G., Derzsy N. *Scaling in income inequalities and its dynamical origin*. *Physica A*, Vol. 549, 124491, (2020)
- [15] Biró T.S., Néda Z. *Unidirectional random growth with resetting*. *Physica A*, Vol. 499, 335-361 (2018)
- [16] Pap I., Varga L., Afifi M., Gere I., Néda Z. *Scaling in the space-time of the Internet*. *Scientific Reports*, 9:9734, (2019)
- [17] Abul-Magd A. Y. *Wealth distribution in an ancient Egyptian society*. *Physical Review E*, Vol. 66, 057104, (2002)

- [18] Hegyi G., NÉda Z., Santos M.A. *Wealth distribution and Pareto's law in the Hungarian medieval society*. Physica A: Statistical Mechanics and its Applications Vol. 380, 271-277, (2007)
- [19] Definition of the word *income*, Cambridge Dictionary, Source: <https://dictionary.cambridge.org/dictionary/english/income>
- [20] Definition of the word: *wealth*, Cambridge Dictionary, Source: <https://dictionary.cambridge.org/dictionary/english/wealth>
- [21] Davies J. B., Shorrocks A. F. *Chapter 11 The distribution of wealth*. Handbook of Income Distribution, Elsevier, Vol. 1, 605-675, (2000)
- [22] Derzsy N., NÉda Z., Santos M.A. *Income distribution patterns from a complete social security database*. Physica A: Statistical Mechanics and its Applications Vol. 391, 5611-5619. (2012)
- [23] Data sources for income, United States of America:
Current population survey, 2014 annual social and economic supplement/income distribution to \$250, 000 or more for males and females: 2013, 2014. PINC-11.-table Income Distribution to \$250,000 or More for Males and Females: 2013, U.S. Census Bureau, (2013), Source: <https://www.census.gov/data/tables/time-series/demo/income-poverty/cps-pinc/pinc-11.2013.html>
- [24] Data sources for income, Finland:
Finland's PxWeb databases/11bq – Number of income recipients, income estimation, deductions and taxes by income-class, Statistics Finland's free-of-charge statistical databases, (2017), Source: http://pxnet2.stat.fi/PXWeb/pxweb/en/StatFin/StatFin__tul__tvt/statfin_tvt_pxt_11bq.px/
- [25] Data sources for income, Russia:
Russia in Figures, Statistical Handbook, M. Rosstat (Ed.), (2017), Source: https://www.gks.ru/free_doc/doc_2017/rusfig/rus17e.pdf
- [26] Data sources for income, Australia:
Taxation statistics 2010-11, Table 14: Individuals tax, Percentile distribution, by taxable income, 2010-11 income year, Australian Taxation Office Australian Government, (2012), Source: https://www.ato.gov.au/uploadedFiles/Content/CR/Research_and_statistics/In_detail/Downloads/cor00345977_2011IND14.xlsx
- [27] Data sources for income, Japan:
The 141th National Annual Statistics Report FY2015, National Tax Agency, Japan, Source: <https://www.nta.go.jp/publication/statistics/kokuzeicho/h27/h27.pdf>
The 142th National Annual Statistics Report FY2016, National Tax Agency, Japan, Source: <https://www.nta.go.jp/publication/statistics/kokuzeicho/h28/h28.pdf>
The 143th National Annual Statistics Report FY2017, National Tax Agency, Japan, Source: <https://www.nta.go.jp/publication/statistics/kokuzeicho/h29/h29.pdf>
- [28] Data sources for wealth, United States of America, Russia, France:
World inequality database, data repository, 2021
Source:
<https://wid.world/data/> (Accessed 23 February 2021).
- [29] Atkinson A.B., Harrison A.J. *Distribution of Wealth in Britain* Cambridge Univ. Press, New York, (1978)

- [30] Drăgulescu A., Yakovenko V. *Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States*. Physica A: Statistical Mechanics and its Applications, Vol. 299, 213-221, (2001)
- [31] Chakrabarti B.K., Chakraborti A., Chatterjee A. *Econophysics and Sociophysics: Trends and Perspectives*. Wiley-VCH Verlag GmbH & Co. KGaA, (2006)
- [32] Dragulescu A., Yakovenko V. *Evidence for the exponential distribution of income in the USA*. European Physical Journal, B 20, 585-589, (2001).
- [33] Yakovenko V., and Rosser J. B., *Colloquium: Statistical mechanics of money, wealth, and income*. Reviews of Modern Physics, Vol. 81, 1703-1725, (2009)
- [34] *Data repository for the Forbes Rich list, years 2001-2019*. Source : <http://datahub.io/gavram/top100-richest-people>
- [35] Bouchaud J.P., Mézard M. *Wealth condensation in a simple model of economy*. Physica A, Vol. 282, 536-545, (2000)
- [36] Biró T.S., Néda Z., Telcs A. *Entropic divergence and entropy related to nonlinear master equations*. Entropy Vol. 21, 993, (2019)
- [37] Alvaredo F., Atkinson A., Chancel L., Piketty T., Saez E., Zucman G. *Distributional National Accounts Guidelines: Methods and Concepts Used in the World Inequality Database*. WID working paper, (2020)
- [38] Gini C. *On the measure of Concentration with Special Reference to Income and Statistics*. Colorado College Publication, General Series No. 208, 73-79, (1936)
- [39] Lorenz M.O., *Methods of measuring the concentration of wealth*. Publications of the American Statistical Association Vol. 9, 209-219, (1905)
- [40] Amiri S., Brian R. *Internet penetration and its correlation to gross domestic product: An analysis of the nordic countries*. International Journal of Business, Humanities and Technology Vol. 3, 50-60, (2013)
- [41] *Definition of the Internet according to the Encyclopedia Britannica online*.
Source:
<https://www.britannica.com/technology/computer/One-interconnected-world>
- [42] *Computer network communication protocol layers according to the Open System Interconnection (OSI) model*.
Source:
<https://www.cloudflare.com/learning/ddos/glossary/open-systems-interconnection-model-osi/>
- [43] Postel J., *Internet Control Message Protocol*, STD 5, RFC 792, September 1981
- [44] *Linux manual page for the "ping" command*.
Source: <https://linux.die.net/man/8/ping>
- [45] Augustin B., Cuvellier X., Orgogozo B., Viger F., Friedman T., et al. *Avoiding traceroute anomalies with Paris traceroute*. IMC 2006 - 6th ACM Internet Measurement Conference, Oct 2006, Rio de Janeiro, Brazil. pp.153-158, (2006)
- [46] Internet topology dataset: *The CAIDA IPv4 Routed/24 Topology Dataset – 2017*,
Source:
https://www.caida.org/catalog/datasets/ipv4_routed_24_topology_dataset/
- [47] Mahadevan P. et al. *The internet as-level topology: Three data sources and one definitive metric*. ACM SIGCOMM Computer Communication Review (CCR), Vol. 36, Issue 1, 17-26, (2006)

-
- [48] Yook S., Jeong H., Barabasi A. *Modeling the Internet's large-scale topology*. Proceedings of the National Academy of Sciences of the United States of America Vol. 99, 13382–13386, (2002)
- [49] Caesar M., Rexford J., *BGP Routing Policies in ISP Networks*. IEEE Network, Vol. 19, (2005)
- [50] Zhang-Shen R., McKeown N. *Designing a predictable Internet backbone network*. Proceedings of Third Workshop on Hot Topics in Networks (HotNets-III), Vol. 10, (2004)
- [51] Isaac Asimov, *Foundation* Doubleday, (1951)