



UNIVERSITATEA BABEȘ-BOLYAI
BABEȘ-BOLYAI TUDOMÁNYEGYETEM
BABEȘ-BOLYAI UNIVERSITÄT
TRADITIO ET EXCELLENTIA

Dinamica sistemelor sociale: o abordare din perspectiva fizicii computaționale

DYNAMICS IN SOCIAL SYSTEMS:
A COMPUTATIONAL PHYSICS APPROACH

Rezumatul tezei de doctorat

TEZĂ DE DOCTORAT

DE

LEVENTE VARGA

LA

DEPARTAMENTUL DE FIZICĂ

CONDUCĂTOR ȘTIINȚIFIC

PROFESOR DR. ZOLTÁN NÉDA

UNIVERSITATEA BABEȘ-BOLYAI

CLUJ-NAPOCA, ROMÂNIA

2019

Dinamica sistemelor sociale: o abordare din perspectiva fizicii computaționale

DYNAMICS IN SOCIAL SYSTEMS:
A COMPUTATIONAL PHYSICS APPROACH

ABSTRACT

În cadrul tezei investigăm sisteme sociale folosind o gamă largă de instrumente și metode ale fizicii computaționale și a fizicii statistice. Prezentăm studii ale teoriei jocurilor evolutive, mobilității umane și rețele de citare cu scopul de a găsi universalități în dinamica sistemelor sociale complexe.

După o introducere generală, în Capitolul 1 studiem distribuțiile de strategii în cazul jocurilor spațiale evolutive cu mai mulți participanți. Pe parcursul acestor jocuri jucătorii dispuși pe diferite tipuri de rețele își schimbă strategiile folosind reguli de actualizare a strategiei de imitare sau logit. Jucătorii pot interacționa cu un anumit număr de vecini, ceea ce duce la cooperare, dezertare sau invazie în cadrul sistemului investigat.

În Capitolul 2 analizăm legile universale ale fluxurilor de navetiști și mobilității umane. Examinăm diferite rețele de transport aerian și de navetă din Ungaria, Italia, Europa, SUA și din întreaga lume. Se studiază viteza medie aparentă în funcție de distanța de deplasare prelucrând date experimentale obținute pe rețele de transport. De asemenea, studiem distribuția fluxurilor de navetiști în funcție de densitatea populației, utilizând seturi de date de recensământ și informații privind locurile de muncă. Vom folosi modele originale de tip radiație pentru a evalua și a explica rezultatele obținute.

În Capitolul 3 studiind probleme de scientometrie argumentăm necesitatea normalizării în cazul comparării producției științifice în diferite domenii de cercetare. Indicatorii individuali ai articolelor științifice pot fi diferiți pentru o gamă largă de discipline, cum ar fi fizica și matematica, unde numărul citărilor este foarte diferit. Studiind benchmark-uri și rețele de publicare, prezentăm câteva metode de grupare care ajută la separarea domeniilor științifice. De asemenea, prezentăm rezultate interesante privind legi de scalare universale în cazul publicațiilor științifice și în cazul postărilor de pe Facebook.

Cuvinte cheie : universalități, scalare, sisteme sociale, sisteme complexe, dinamică, fizica computațională, fizica statistică, jocurile evolutive, mobilitatea umană, modele de navetă, rețele, rețele de citare, rețele sociale

Cuprinsul tezei de doctorat

INTRODUCERE	1
Jocuri evolutive	1
Mobilitate umană	2
Dinamica citărilor și rețele complexe	3
1 JOCURI EVOLUTIVE	5
1.1 Prezentare generală	5
1.2 Analiza distribuției strategiilor în jocul de potrivire a monedelor cu două tipuri de jucători pe o rețea pătratică și pe rețele bipartit	8
1.3 Structuri auto-organizate, invazii și speciații într-un joc evolutiv de piatră-hârtie-foarfecă, folosind actualizări strategice sincronizate și stohastice	14
1.4 Componentele de câștig și efectele lor într-un joc spațial evolutiv al dilemei prizonierului cu trei strategii pe o rețea pătratică	20
1.5 Procesele de invazie anizotropă obținute prin schimbarea tăriei componentelor autodependente și interdependente într-un joc evolutiv cu două strategii pe o rețea pătratică și cu regulă de imitație	25
1.6 Concluzie	30
2 MOBILITATEA UMANĂ	33
2.1 Prezentare generală	33
2.2 Cu cât mai departe călătorim, cu atât mai repede mergem. Scalarea netrivială a vitezei medii în funcție de distanța de călătorie.	35
2.3 Modelul de radiație, modelul de radiație cu selecție și un model de radiație optimizat pe costurile de deplasare. Aplicabilitatea acestor modele pentru a descrie fluxurile de navetiști în Ungaria.	42
2.4 Modelul de radiație, modelul de gravitație și modelul de “flow and jump” pentru studierea fluxurilor de navetiști în trei regiuni geografice diferite	51
2.5 Concluzie	59
3 DINAMICA CITĂRILOR ȘI REȚELE COMPLEXE	61
3.1 Prezentare generală	61
3.2 O abordare geometrică pentru detectarea comunităților bazate pe diagramele Voronoi	63
3.3 Clusterizare stohastică bazată pe partajare de tip Voronoi și identificarea comunităților	69
3.4 Normalizarea indicatorilor scientometrici ai publicațiilor individuale folosind metode locale de detectare a clusterelor bazate pe proprietățile structurale ale benchmark-uri și a rețelilor de citare	74
3.5 Publicații științifice și Facebook: aceeași lege de popularitate! – distribuția citărilor pentru publicații și distribuția “share”-urilor pentru posturile de pe Facebook	82
3.6 Concluzie	85

4	CONCLUZIE GENERALĂ	87
5	PUBLICAȚII LEGATE DE TEZĂ	91
5.1	Articole științifice	91
5.2	Articole apărute în proceedings-uri ale unor conferințe	92
5.3	Prezentări la conferințe	92
5.4	Postere prezentate la conferință	92
	BIBLIOGRAFIE	93

Cuprins

INTRODUCERE	1
1 JOCURI EVOLUTIVE	3
2 MOBILITATEA UMANĂ	10
3 DINAMICA CITĂRILOR ȘI REȚELE COMPLEXE	19
4 CONCLUZII	26
5 PUBLICAȚII LEGATE DE TEZĂ	28
5.1 Articole științifice	28
5.2 Articole apărute în proceedings-uri ale unor conferințe	28
5.3 Prezentări la conferințe	29
5.4 Postere prezentate la conferințe	29
BIBLIOGRAFIE SELECTATĂ	30

Introducere

Sistemul social – după definiția lui Parsons (savantul care a definit conceptul de sistem în sociologia modernă) – este o structură ordonată, cu o interdependență între părți, în care fiecare parte are o localizare determinată și un rol determinat în cooperare [1]. Exemple de sisteme sociale sunt familiile, comunitățile, națiunile, industriile, care depind de diferite caracteristici comune, cum ar fi locația, normele culturale, religia, statutul socio-economic, etc.

Sistemele sociale sunt exemple cunoscute de sisteme complexe, oferând o multitudine de domenii de cercetare interesante. Cercetătorii au la dispoziție foarte multe seturi de date în format electronic pe care pot să-și testeze modelele concepute și să dezvăluie eventualele legi de universalitate [2]. În aceste cercetări, procesăm seturile de date disponibile, simulăm datele experimentale și dezvoltăm modele pentru identificarea și studierea comportamentului dinamic ale sistemelor. Domeniile principale ale fizicii utilizate în astfel de studii sunt fizica computațională împreună cu teoria jocurilor evolutive, fizica statistică și analiza rețelelor complexe. În aceasta teză aducem contribuții în acest sens, studiind fenomene precum cooperarea, concurența, mobilități umane și cluzterizarea.

Originalitatea tezei constă în faptul că oferim modele noi, abordări și metode de analiză a datelor în domeniul dinamicii sistemelor sociale. În teză se prezintă rezultatele științifice obținute, grupându-le în trei domenii mari.

Primul domeniu este teoria jocurilor evolutive, care este o aplicație a teoriei matematice a jocurilor prin plasarea participanților într-un context biologic [3]. Jucătorii imită unele reguli evolutive în ceea ce privește selecția darwiniană sau modifică strategiile lor după raționamentul individual al sistemelor sociale. Jucătorii sunt plasați în mod tipic pe o grilă sau o rețea și interacționează cu un anumit număr de vecini.

Metodele fizicii statistice sunt aplicate direct în jocurile evolutive, de exemplu, modelul Ising bidimensional într-un câmp magnetic cu spini sus și jos [4]: stările de spin reprezintă strategiile jucătorilor, și dinamica Glauber permite întoarcerea spinilor (strategiilor) cu o anumită probabilitate [5]. Probabilitățile sunt calculate pentru regulile de actualizare a strategiei de imitare sau logit, iar în timpul jocurilor jucătorii încearcă să-și maximizeze câștigul.

Un alt element important al fizicii statistice în jocurile evolutive este simularea Monte Carlo [6]. Folosind aceasta metodă, putem simula pe calculator jocurile evolutive.

În această teză sistemele sociale modelate de teoria jocurilor evolutive sunt examinate prin măsurarea distribuțiilor diferitelor strategii. Putem astfel observa formele domeniilor de cooperare sau de concurență a jucătorilor, de exemplu ale acelora, care urmăresc același scop strategic.

Al doilea domeniu este mobilitatea umană, care este importantă în multe sisteme sociale. Dacă ne gândim la traficul zilnic sau chiar la dezvoltarea economică și urbană, este important să înțelegem legile universale care le guvernează [7]. Călătoriile de zi cu zi ale indivizilor sunt semnificative, așa cum este prezentat în studiul lui González și colab. [8]. Indivizii se întorc adesea la câteva locații foarte frecventate,

iar traiectoriile lor arată o bună predictibilitate.

Seturi de date electronice care sunt foarte utile în studiul mobilității umane sunt seturile de date de recensământ privind rutele de trafic, densitatea populației, rețelele rutiere și aeriene, precum și timpii de călătorie și întârzierile cauzate de nodurile de trafic. Aceste date pot fi analizate prin metodele fizicii statistice și înțelese în cadrul diferitelor modele: modelul de gravitație [9], modelul de radiație și modelul de radiație cu selecție [10, 11].

În teză sunt date două generalizări originale ale modelului clasic de radiație: unul este modelul de radiație optimizat pentru costurile de călătorie, în care pentru alegerea locurilor de muncă se ia în considerare și costul de transport pentru navetă [12]. Celălalt este modelul de “flow and jump”, care se bazează pe o ecuație de tip master. [13].

Al treilea domeniu se referă la reprezentările prin rețea ale sistemelor complexe, în special ale sistemelor sociale, metodă des folosită în științele sociale, în teoria informațiilor, în neurologie și biologie, etc. [14, 15]. Reprezentarea conexiunilor dintre elemente prin legăturile dintre nodurile unei rețele este o vizualizare utilă pentru relațiile structurale ale oricărui sistem complex.

Pentru a înțelege lumea complexă a relațiilor științifice, se poate folosi atât rețelele de citare, colaborare cât și rețelele sociale. Abordările fizicii computaționale împreună cu metodele clasice de procesare a rețelilor, cum ar fi vizualizarea, clusterizarea sau distribuțiile statistice, ajută la analiza caracteristicilor structurale ale rețelilor [15]. Identificarea comunităților este importantă pentru o prelucrare mai ușoară și o mai bună înțelegere a topologiei rețelilor complexe.

Programarea pe calculator combinată cu metodele fizicii statistice oferă o varietate de cadre, instrumente software și algoritmi pentru analiza rețelilor. Folosind aceste instrumente, putem identifica comunități, subrețele și structuri pentru benchmarking și pentru compararea rețelilor din lumea reală.

Există multe modalități de a compara comunitățile create prin clusterizarea rețelilor. Un exemplu de comparare este normalizarea “cross-field” a indicatorilor scientometrici [16–18]. Distribuția indicatorilor studiați din fiecare domeniu științific poate fi normalizată folosind o funcție de putere [19]. După normalizare se poate face o comparație pe aceeași scală între indicatorii calculați pentru două sau mai multe domenii diferite, cum ar fi fizica și matematica, unde indicatorii respectivi pot fi echivalați printr-o constantă de proporționalitate. Tot în cadrul acestui capitol se studiază universalități și similitudini observate în distribuția share-urilor pe Facebook și a citărilor în știința. Se arată universalitatea unei distribuții de tip Tsallis-Pareto care se poate explica tot cu ajutorului unei ecuații de tip masters [20].

Teza este organizată în trei capitole extinse corespunzătoare domeniilor mai sus menționate. Capitolul 1 cuprinde simulările Monte Carlo, calculele teoretice și rezultatele obținute pentru jocurile evolutive. În Capitolul 2 prezentăm analiza datelor privind mobilitatea umană și fluxurile de navetiști. Aici discutăm despre aplicabilitatea unor modele simple în reproducerea datelor experimentale. În Capitolul 3 se prezintă rezultate pentru detectarea comunităților, despre normalizarea indicatorilor științifici individuali și despre similitudinea dintre publicațiile științifice și posturile de pe Facebook. La sfârșitul tezei, oferim o sinteză scurtă a rezultatelor obținute.

“Teoria jocurilor evolutive este o modalitate de a gândi despre evoluție la nivel fenotipic atunci când starea de fitness a fenotipurilor particulare depinde de proporția lor în populație.”

John Maynard Smith

1

Jocuri evolutive

În lucrare se studiază jocurile evolutive spațiale cu participanți multipli. Pe aceste sisteme, jucătorii sunt plasați pe o grilă sau o rețea și interacționează cu un anumit număr de vecini, așa cum se arată în Figura 1.1. În timpul unui joc, în fiecare etapă, indivizii participă la toate jocurile potențiale, alegând o strategie posibilă. Interacțiunea acestora cu vecinii selectați este reprezentată de matricea de câștig:

$$\mathbf{A} = \begin{pmatrix} +1 & -1 \\ -1 & +1 \end{pmatrix} \quad (1.1)$$

În exemplul de mai sus (1.1), matricea $\mathbf{A}$ indică câștigul jucătorilor dacă selectează o combinație dintre strategiile posibile. Jocurile evolutive pot fi deterministice sau stocastice cu regulă de actualizare dinamică secvențială sincronizată sau aleatorie, în funcție de alegerea strategiei de imitație sau a celei oarbe pentru a obține un câștig individual mai mare.

Metoda ajustării prin imitație este utilizată în mod analog cu procesele biologice, atunci când jucătorii își pot imita vecinii mai de succes. Regula de ajustare a strategiei oarbe este așa-numita regulă logit din economie și modele de fizică [21, 22]. În această situație fiecare jucător își poate calcula câștigul și preferă să aleagă o strategie cu câștig mai mare, în funcție de valorile așteptate.

Pentru a analiza jocurile evolutive, cercetătorii consideră foarte des simulări de tip Monte Carlo (MC) pe rețele cu N noduri. În majoritatea cazurilor simularea pornește de la o condiție inițială aleatorie, dar, condiția inițială poate fi și un tipar predefinit. În timpul simulărilor, după o perioadă de termalizare adecvată t_{th} , începem măsurarea datelor statistice pe un timp de eșanționare t_s . Aceste valori sunt ajustate la parametrii sistemului și, de obicei, folosim condiții de limită periodice.

În timpul simulărilor MC, am folosit metodele dinamice de cluster sau aproximările câmpului generalizat [5]. Pentru aceste metode, calculăm toate probabilitățile de configurații astfel încât rezolvăm numeric un set de ecuații de mișcări și, ca rezultat, putem da predicții analitice pentru cantitățile studiate.

Folosind metodele de mai sus, studiem câteva întrebări din teoria jocurilor evolutive și discutăm rezultatele obținute. Prima problemă studiată este jocul de potrivire a monedelor pentru două tipuri de rețele. Considerăm apoi un joc evolutiv de piatră-hârtie-foarfecă cu ajustarea sincronizată a strategiei. De asemenea, discutăm relația dintre strategiile unui joc evolutiv spațial extins de tipul dilema prizonierului. Aici, strategia introdusă pentru câștig-staționare-pierdere-schimbare evaluează rezultatul ultimei runde,

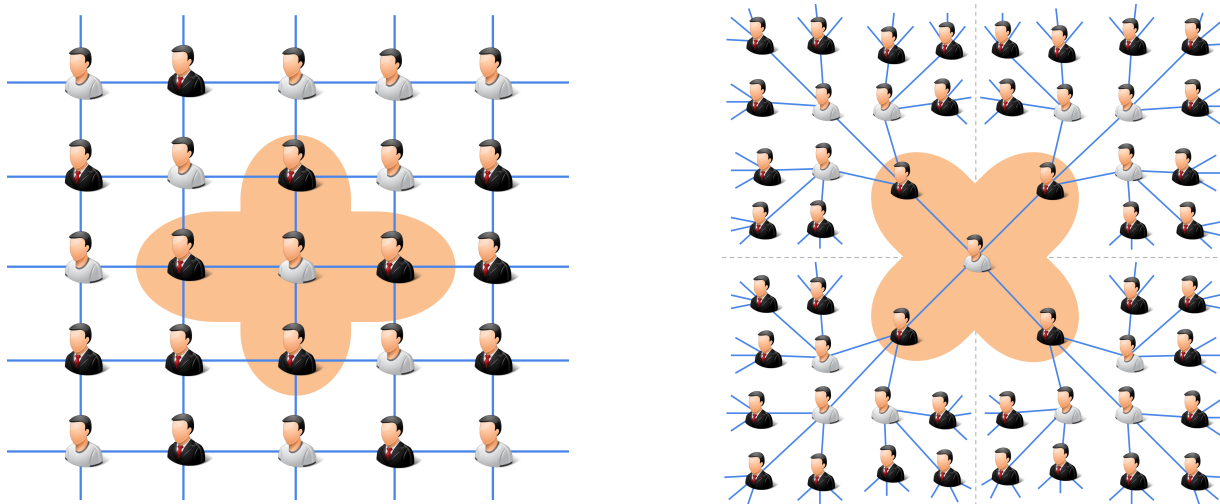


Figure 1.1: Jucători evoluționiști ai jocurilor cu vecinii care interacționează pe o grilă pătrată bidimensională (stânga) și pe o grilă Bethe fără buclă (dreapta).

iar prin schimbarea strategiilor, jucătorii pot mări câștigul. În cele din urmă, am studiat procesele de invazie pe o grilă pătrată în cazul jocurilor evolutive cu două strategii.

Mai specific: în **primul** rând, analizăm un joc evolutiv, cunoscut ca jocul de potrivire a monedelor, unde două tipuri de jucători, X și Y , se află pe rețele bipartite [23]. Studiem aceste jocuri evolutive pe o grilă pătrată și pe o grilă Bethe bipartită fără buclă, așa cum se vede în Figura 1.1. În ambele cazuri, fiecare jucător de tip X interacționează cu cei patru vecini conectați de tipul Y , similar cu pătratele albe și negre pe o tablă de șah. Pe subfigura din stânga din Figura 1.1 este vizibilă o distribuție aleatorie a jucătorilor în ceea ce privește tipul de jucători. Pe subfigura din dreapta a Figurii 1.1 fiecare jucător de tip X este conectat la cei patru vecini de tip Y (jucătorul alb din centru este conectat cu patru jucători negri învecinați).

Folosind aceste rețele și cele două tipuri de jucători, examinăm efectele interacțiunilor de potrivire a monedelor, iar apoi studiem distribuția spațială a celor două opțiuni, numite cap (H) și pajură (T). În jocul tradițional de potrivire a monedelor, cei doi jucători cad de acord în primul rând asupra câștigătorului: dacă fețele monedelor sunt identice sau diferite. Apoi ascund o monedă în palme cu capul sau pajura în sus și își dezvăluie alegerile simultan. Câștigătorul primește moneda adversarului. Jucătorul de tip X câștigă cu opțiuni strategice (H,H) sau (T,T) , iar jucătorul de tip Y câștigă cu combinații strategice (H,T) sau (T,H) .

Prin urmare, studiem distribuția strategiei prin efectul jocului de potrivire a monedelor pentru două tipuri de rețele atunci când variază nivelul de zgomot în regula dinamică de tip Glauber [24]. După cum se vede în exemplul din Figura 1.2, ilustrăm prezența diferitelor strategii prin reprezentarea lor prin cutii albe (H) și negre (T) de pe figură.

Sistemele studiate sunt analizate prin efectuarea de simulări MC pe grilă pătratică de $N = L \times L$ noduri, cu condiții limită periodice și rețele bipartite aleatorii regulate cu N noduri. În timpul simulărilor, N este variat de la $N = 2.5 \times 10^5$ până la 4×10^6 . Simularea a pornit de la o stare inițială aleatoare, iar după un timp de termalizare t_{th} , datele statistice sunt obținute prin medierea pe un timp de eșantionare t_s . Această valoare t_{th} se modifică de obicei între 2 000 și 10^6 pași Monte Carlo (MCS) iar valoarea t_s este



Figure 1.2: Distribuția strategiei în cazul jocului de potrivire a monedelor. Imaginea de 50×50 prezintă o distribuție de strategii reprezentat cu boxe de culoare albă (H) și negru (T).

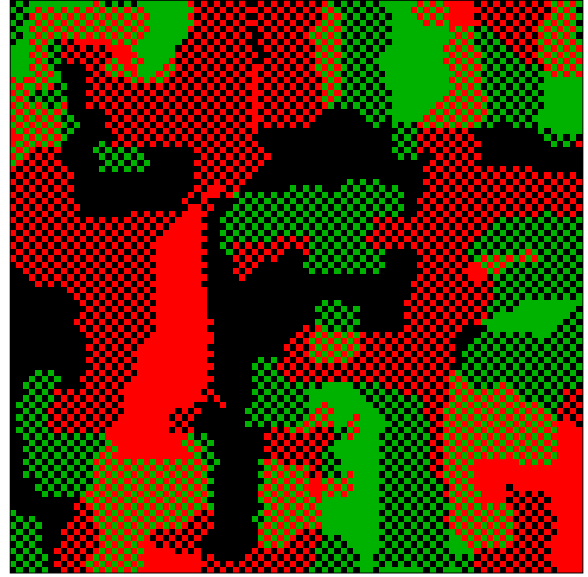


Figure 1.3: O stare instantanee tipică pentru distribuția spațială a strategiilor într-un bloc de 90×90 noduri cu o grilă pătrată și la nivel de zgomot $K = 0.4$. Culoarea neagră reprezintă strategia R, roșu reprezintă strategia P, și culoarea verde indică strategia S.

modificată de la 50 000 la 3×10^6 MCS, unde în cadrul unității de timp (pas Monte Carlo) fiecare nod are în medie o șansă de a-și modifica starea.

Simulările MC au confirmat cantitativ absența corelațiilor dintre cele mai apropiate locații vecine. Cu toate acestea, am constatat slabe corelații negative între vecinii de gradul al doilea și al treilea.

După rezultatele de simulare, analizăm predicțiile teoretice. Din aceste rezultate calculate s-a determinat valorile funcției de corelație. Rezultatele analitice s-au comparat cu rezultatele simulărilor MC.

Studiile au fost repetate pe rețele obișnuite bipartite, pentru același număr de vecini, $k = 4$. Acest lucru este necesar pentru analiza caracteristicile topologice ale rețelei de conectivitate. Cele mai impresionante diferențe sunt magnitudinile scăzute ale corelațiilor pe rețele bipartite aleatorii obișnuite și absența lor pentru nodurile vecine de gradul trei.

Ca și o a doua problemă s-a studiat un joc evolutiv spațial piatră-hârtie-foarfecă (RPS) pe o grilă pătrată în care jucătorul obține câștigul din jocurile cu cei patru vecini ai săi [25]. Aceștia utilizează ajustarea sincronizată a strategiei în conformitate cu regula logit. Pentru jocul RPS tradițional, jucătorii aleg simultan una din cele trei strategii numite piatră (R), hârtie (P) și foarfecă (S). Conform regulilor jocului, o strategie este superioară celei de-a doua și inferioară celei de-a treia: piatra bate foarfeca, foarfeca bate hârtia și hârtia bate piatra, adică jocul reprezintă dominarea ciclică a strategiilor.

Sistemele cu interacțiune RPS arată o oscilație globală în rata de apariție a strategiilor. În general, într-o populație structurată, atunci când jucătorii sunt plasați într-o rețea, oscilația evoluează spre un ciclu limitat sau crește până când o singură strategie rămâne în viață. În cazul bidimensional, simulările numerice arată supraviețuirea tuturor celor trei strategii prin tiparuri auto-organizate.

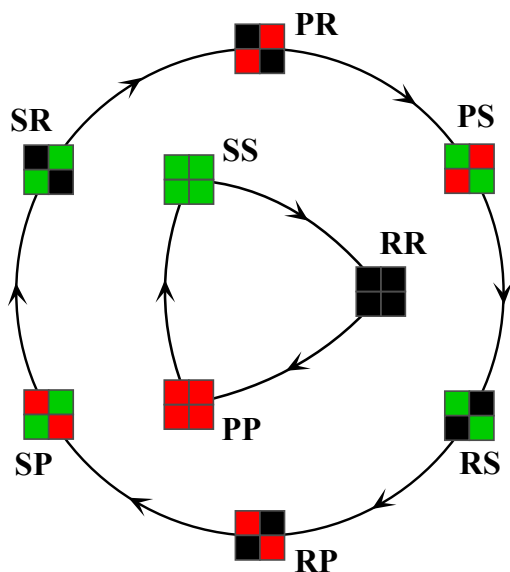


Figure 1.4: Diagrama de curgere pentru comportamentul ciclic al domeniilor. Liniile solide (negre) arată modul în care compoziția domeniului se schimbă în fiecare generație, datorită regulii de adoptare a strategiei logit. Codul de culoare este același ca în Figura 1.3. Tipurile de domenii sunt notate folosind numele strategiilor corespunzătoare. Este important de a reține diferite nume pentru structurile de șah și structura anti-șah. Domeniile pure revin la aceeași stare în trei generații, asociațiile de tip șah realizează acest lucru în șase generații. Interesant este faptul că cele nouă asociații strategice disponibile formează două cicluri disjuncte în loc de un ciclu complet.

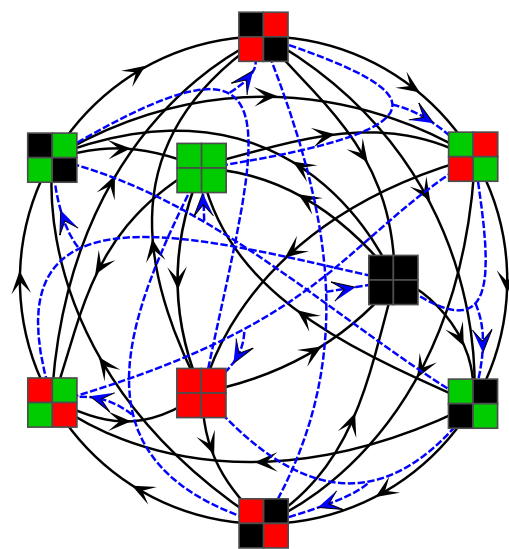


Figure 1.5: Diagrama de curgere pentru procesele de invazie și de speciație între diferitele asociații de strategii (domenii). Săgețile de pe linia solidă (neagră) arată asociația câștigătoare atunci când cele două asociații conectate de linia neagră se întâlnesc reciproc. Linii punctate (albastre) afișează cazurile în care întâlnirea a două asociații are ca rezultat apariția unei a treia specii. Săgeata indică spre asocierea nou creată. Semnificația culorilor de strategie este aceeași ca în Figura 1.3.

Folosind oscilațiile acestui sistem în cadrul acestor domenii de strategii, am studiat o relație competitivă de nouă specii, în care există trei strategii pure R, P, S și șase alte specii mixte de strategie, cu o structură de tip șah, în care structura alb-negru de șah este ocupată de două strategii diferite, după cum se poate vedea în Figura 1.4. Jucătorii sunt situați pe o grilă pătrată cu condiții de limită periodice și își colectează câștigurile din jocurile RPS cu vecinii cei mai apropiați.

Am studiat modelul efectuând simulări MC pe grile pătrate de dimensiuni $N = 4 \times 10^4$ până la 10^6 . De obicei, simularea a pornit dintr-o stare de strategie aleatorie, dar în cazul invaziilor am aplicat și condiții inițiale predefinite. De asemenea, am folosit un timp de termalizare $t_{th} = 1\,000 - 5\,000$ MCS pentru stabilizarea inițială a sistemului. Apoi, în timpul de eșantionare $t_s = 10^4 - 10^6$ MCS, am măsurat datele relevante.

Plecând dintr-o stare inițială aleatorie, am observat un proces de formare de clustere care indică o oscilație sincronă la niveluri scăzute de zgomot. În timpul procesului de creștere a domeniului, în sistem se creează un tipar auto-organizat. Un exemplu de stare instantanee este prezentat în Figura 1.3, unde se pot distinge toate cele nouă tipuri de strategii.

În urma regulilor de ajustare, jucătorii aleg strategia care le oferă cele mai mari câștiguri. În primul rând, luăm în considerare domeniile omogene, care sunt notate cu aceleași perechi de strategii RR, PP și SS, poziționate pe cele două subrețele. Aici putem observa un comportament ciclic, de exemplu, în momentul $t = 0$ toți jucătorii folosesc strategia R, apoi în următorul, $t = 1$, toți vor alege P, și după aceea, în $t = 2$, toate strategiile vor fi S. Aceste modificări ciclice sunt ilustrate în inelul interior din Figura 1.4.

În al doilea caz, perechile pot fi diferite. De exemplu, strategiile mixte RP denotă o dispersie uniformă a strategiilor R și P pe cele două subrețele. În acest domeniu, jucătorii P nu doresc să-și schimbe strategiile, dar jucătorii R nemulțumiți sunt forțați să-și modifice strategiile la S. În consecință, sistemul evoluează în starea SP. Pentru fiecare caz acest comportament ciclic mixt este asemănător celui din inelul exterior din Figura 1.4.

În funcție de valoarea zgomotului K , distingem două cazuri. La niveluri scăzute de zgomot, rezultatele simulării arată că defectele de punct rămân în interiorul tiparurilor ciclice omogene și mixte. Pe măsura creșterii valorii lui K , la niveluri ridicate de zgomot, putem observa mici insule ale altor stări în cadrul domeniilor conturate. Această structură relativ complexă de perechi de strategii și invazii este vizibilă în Figura 1.5.

În continuare, să ne uităm la semnificația relațiilor de invazie intercorelată și de speciație între diferitele tipuri de domenii, marcate cu linii întrerupte negre și albastre. Liniile negre cu săgeata pe ele indică direcția acestor invazii. De exemplu, vedem că perechea de strategii RS invadează perechea RP. Liniile întrerupte albastre cu săgeți reprezintă coliziunea a două domenii de strategie, care dă naștere unei a treia și invadează perechile strategice inițiale.

Rezumând cele două cazuri de mai sus, putem menționa un al treilea caz, când a treia pereche de strategii, cea emergentă nu invadează nici una dintre strategiile inițiale. În această situație, noua pereche de strategii joacă rolul de catalizator și trece prin sistem într-un rol invadator-apărător, dar, în cele din urmă, domeniul superior îl consumă și pe acesta.

În a treia problemă, am încercat să prezentăm evoluția cooperării cu ajutorul unui joc evolutiv al dilemei prizonierului cu trei strategii [26]. În mod tradițional, pe parcursul jocului de dilema prizonierului (PD), doi jucători participanți aleg între strategiile de cooperare (C) și trădare (D). În mod similar, în modelul nostru, jucătorii pot adopta trei strategii: cooperează întotdeauna (AllC), trădează întotdeauna (AllD) și strategia nou introdusă câștig-staționare-pierdere-schimbare (WSLS). WSLS cooperează în primul tur, iar ulterior evaluează rezultatul ultimelor runde. Își schimbă strategia dacă câștigul rezultat este mai mic decât un prag predefinit.

Într-un singur joc, jucătorii selectează între cooperare (C) și trădare (D) și primesc diferite câștiguri în funcție de deciziile lor simultane. Rezultatul cooperării reciproce este recompensa R , iar prețul trădării este pedeapsa P . Un trădător care exploatează un colaborator câștigă tentația T , în timp ce victima exploatarei primește câștigul fraierului S . PD satisface clasamentul $T > R > P > S$. Conform generalizării, putem fixa răsplata ($R = 1$) și pedeapsa ($P = 0$) rămânând astfel doi parametri de câștig.

Studiem modelul evolutiv spațial pe o grilă pătrată cu patru vecini apropiați. Jucătorii joacă un joc iterativ PD (IPD). Pe lângă jocurile evolutive discutate mai sus, studiem ambele mecanisme de update (imitație și logit) în acest caz.

Pe o rețea pătrată de $N = L \times L$, jucătorii pot folosi strategiile AllD, AllC și WSLS. Cu toate acestea, la strategia WSLS am stabilit un parametru de prietenie w , cu valoarea între $R = 1$ și $P = 0$, care definește probabilitatea de cooperare în primul tur; în consecință, valoarea mare w poate fi asociată cu un comportament prietenos. În majoritatea cazurilor, acest parametru este setat la $w = 0.5$, dar analizăm și valori diferite: $w = 0.1$ și 0.9 .

În simulările MC, examinăm modelul pe o grilă pătrată, caracterizată printr-o condiție limită periodică. Dimensiunea utilizată a sistemului este de $N = 200 \times 200 = 40\,000$ limitată de capacitatea noastră de calcul. Într-un pas Monte Carlo, fiecare jucător are opțiunea de a schimba strategia o dată în medie. Pentru diagramele de fază, concentrațiile staționare ale strategiei au fost obținute prin medierea distribuției strategiei. Simularea rulează într-un timp tranzitoriu de $t_{tr} = 50\,000$ MCS, iar medierea se face pentru timpul de eșantionare de $t_s = 10\,000$ MCS.

Folosind relațiile $T > R > P > S$, $R = 1$ și $P = 0$, am studiat analitic caracterul potențial al jocului în cazul $w = 0.5$ prin ajustarea regulii logit. Am studiat interacțiunea și echivalența dintre cele trei strategii AllD, AllC și WSLS.

Am simulat rezultatele teoretice și observăm competiția dintre strategiile AllD și WSLS. Este vizibil faptul că strategia AllC dominată este prezentă la o frecvență extrem de scăzută și practic nu deranjează competiția dintre AllD și WSLS.

De asemenea, am studiat parametrii de prietenie $w \neq 0.5$ pentru $w = 0.1$ și 0.9 . În cazul $w = 0.1$, zona strategiei WSLS neprietenoase s-a lărgit, în plus, în cazul $w = 0.9$, teritoriul strategiei prietenoase WSLS s-a restrâns. Această selecție a parametrilor este ceea ce ne așteptăm dacă vrem să supraviețuim mai bine împotriva strategiei AllD.

Folosind regula imitației ne arată ca rezultatele obținute sunt diferite de rezultatele regulii logit. Unul dintre factorii importanți pe care ar trebui să-i menționăm în cazul imitării este acela, că parametrul w nu mai joacă nici un rol. O diferență semnificativă este faptul că strategia WSLS câștigă teren. Aceste simulări demonstrează impactul pozitiv al imitației în strategia WSLS cooperativă.

Ca o **ultima** problemă, investigăm procesele de invazie prin schimbarea tăriei componentelor autodependente și interdependente [27]. Referindu-ne la întregul spațiu parametric $T - S$, restrângem investigația la regiunea de $T < 1$ și $S < 0$.

Studiem un joc evolutiv cu două strategii, cu regula de imitație. Jucătorii sunt situați pe o grilă pătrată și interacționează cu vecinii cei mai apropiați. Examinăm mișcarea interfeței și vitezele de invazie pentru diferitele faze omogene.

În timpul simulărilor noastre am folosit o grilă pătrată cu $N = L \times L$ noduri, unde jucătorii echivalenți aleg una dintre cele două strategii (1 sau 2). Sistemul are condiții de limită periodice. Jucătorii interacționează cu cei mai apropiați patru vecini.

Am aplicat ajustarea strategiei de imitare cu o mică schimbare. S-au ales doi jucători învecinați și jucătorul selectat adoptă strategia vecinului cu o probabilitate calculată.

Toate simulările noastre sunt inițiate dintr-o stare artificială, în care una dintre strategii formează o insulă circulară în marea celeilaltei strategii. În cazul în care strategia 1 (negru) este marea și strategia 2 (alb) este insula circulară, putem vedea reducerea și dispariția domeniului la interfețele orizontale și

verticale. În situația opusă, dacă cele două strategii schimbă rolurile de mare și de insulă, există și în acest caz un domeniu care se reduce și se micșorează, dar acum, surprinzător, domeniul evoluează de-a lungul interfețelor înclinate și are o contracție mai rapidă.

De asemenea, am studiat comportamentul invaziei atunci când rețeaua pătrată a fost înlocuită cu alte grile bidimensionale. Practic, am observat aceeași atitudine a jucătorilor și strategiilor. Diferențele în orientarea invaziilor au fost observate numai atunci când jucătorii au colectat câștigurile dintr-un joc cu vecinii cei mai apropiați sau cu vecinii de gradul doi.

“Trăim într-o lume în care coexistă mari incompatibilități: scala umană și scala supraumană, stabilitatea și mobilitatea, permanența și schimbarea, identitatea și anonimitatea, inteligibilitatea și universalitatea.”

Kenzo Tange

2

Mobilitatea umană

În mod similar cu alte studii de mobilitate umană, și aici avem de-a face cu înțelegerea și dezvăluirea legilor universale care guvernează mobilitățile umane. Ne-am propus să analizăm diferitele moduri de transport și distribuția fluxurilor de mobilitate între așezări.

În prima parte a studiului analizăm mai multe rețele de transport uman, cum ar fi rețeaua rutieră a Ungariei și rețeaua rutieră interstatală din SUA continentală, zborurile directe între aeroporturile principale din Europa și rețeaua de călătorii aeriene din SUA.

Examinând timpul de călătorie în funcție de distanța de deplasare, studiem variația vitezei de deplasare medie cu distanța de deplasare. Corelăm rezultatele pentru timpul de călătorie cu timpul petrecut în principalele noduri, structura rețelelor de călătorie și limita de viteză a drumurilor și vehiculelor. Investigăm această viteză, folosind atât distanța măsurată pe linia geodezică cât și cea efectiv parcursă în cadrul deplasării.

O altă problemă importantă în domeniul mobilității umane este naveta zilnică. Putem obține rezultate experimentale din datele de recensământ, care conțin multe informații utile, cum ar fi domiciliul și locul de muncă al navetiștilor.

Folosind aceste date și datele de distribuție spațială pentru densitatea populației, investigăm și modelăm fluxurile de navetiști la diferite distanțe. S-au studiat date privind naveta și populația din Ungaria, Italia și SUA și s-au obținut rezultate cu caracter universal pentru fluxurile de mobilitate. S-au comparat modelele existente și cele noi pentru a fi datele fluxurilor de navetiști. Astfel de modele clasice sunt modelul original de Radiație (RM) [10], modelul de Gravitație (GM) [9] și modelul generalizat de Radiație cu Selecție (RMwS) [11] care au fost utilizate cu succes mai demult pentru a modela selecția locului de muncă a navetiștilor. În teză oferim două noi generalizări ale modelului de radiație. Una cu optimizarea costurilor de călătorie și cealaltă cu aplicarea unei ecuații master, care permite atât procesele de salt, cât și cele de flux local.

Ca un **prim** rezultat semnificativ, am descoperit o relație universală între viteza medie a mobilității și distanța de călătorie [28]. Călătoria este o parte integrantă a vieții noastre și știm cu toții că timpul de călătorie nu crește liniar cu distanța de călătorie [7]. Este cunoscut, de exemplu, că uneori este nevoie de mai multe ore pentru a călători la distanțe de câteva sute de kilometri, dar timpul de călătorie nu crește în mod semnificativ în cazul unei călătorii până pe partea opusă a Pământului.

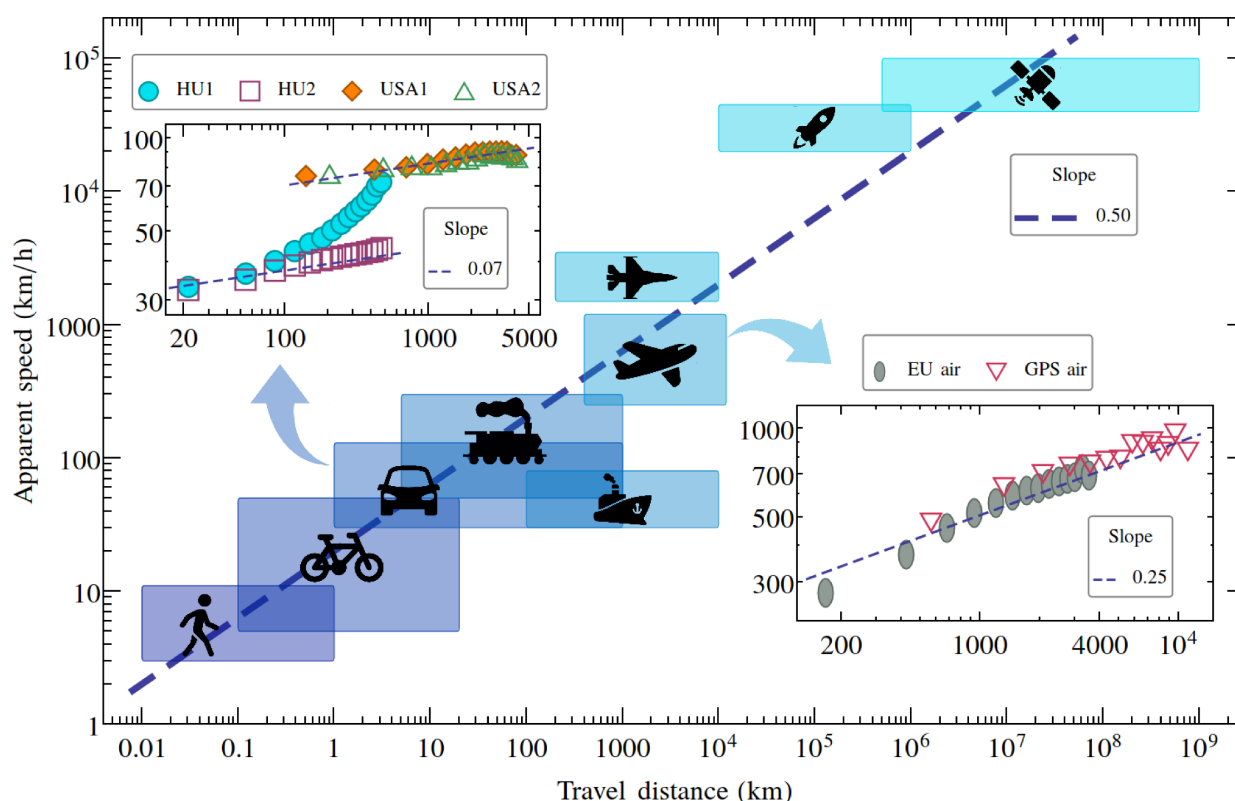


Figure 2.1: Scalele de viteză și de distanță ale călătoriilor umane. Viteza de deplasare aparentă (estimată ca distanța parcursă pe linia geodezică împărțită la timpul de deplasare) ca funcție a distanței de deplasare. Cutiile indică intervale pentru diferite moduri de deplasare. Cele două figuri inserate prezintă câteva rezultate mediate pentru cele două moduri de călătorie care sunt cele mai populare: călătoriile cu mașinile și cu avionul. Liniile punctate din aceste figuri inserate indică tendințele de scalare cu exponentul specificat. Liniile punctate cu pante diferite nu sunt rezultate fitate; ele indică doar tendințe de scalare cu exponenții specificați pentru ghidarea ochilor. Este de observat că axele sunt logaritmice.

Pentru a ilustra această ipoteză de lucru, Figura 2.1 arată scalele estimate pentru distanță și viteză pentru diferite moduri de călătorie începând de la mersul pe jos până la călătorii cosmice. Tendința generală este o lege de putere cu un exponent mediu de 0.5 pe toată scara distanțelor și vitezelor.

În plus, o lege asemănătoare este vizibilă și în inserțiile din Figura 2.1, care reprezintă separat diferitele moduri de transport. În investigațiile noastre, ne concentrăm asupra celor două moduri de călătorie populare, cea cu autoturismul personal și călătoriile aeriene. În ambele cazuri, am obținut aceeași creștere subliniară a timpului de călătorie în funcție de distanța de deplasare, ca în imaginea generală. De exemplu, dacă călătorim de la un oraș la un altul, care este departe, pentru o mare parte din călătorie folosim autostrăzi unde viteza medie este mai mare. Cu toate acestea, atunci când ieșim dintr-un oraș, există multe opriri, blocajele de trafic reducând viteza medie în mod drastic pentru deplasări scurte. Pentru călătoriile cu avionul, distanțele mai mici sunt servite cu avioane mai mici, cu altitudine de călătorie mai mică și viteză de deplasare mai mică, iar viteza medie de deplasare este mult redusă pentru deplasări scurte, datorită manevrelor de decolare, aterizare și parcare.

Așa cum am arătat în cele două figuri inserate din Figura 2.1, am colectat și analizat date referitoare la

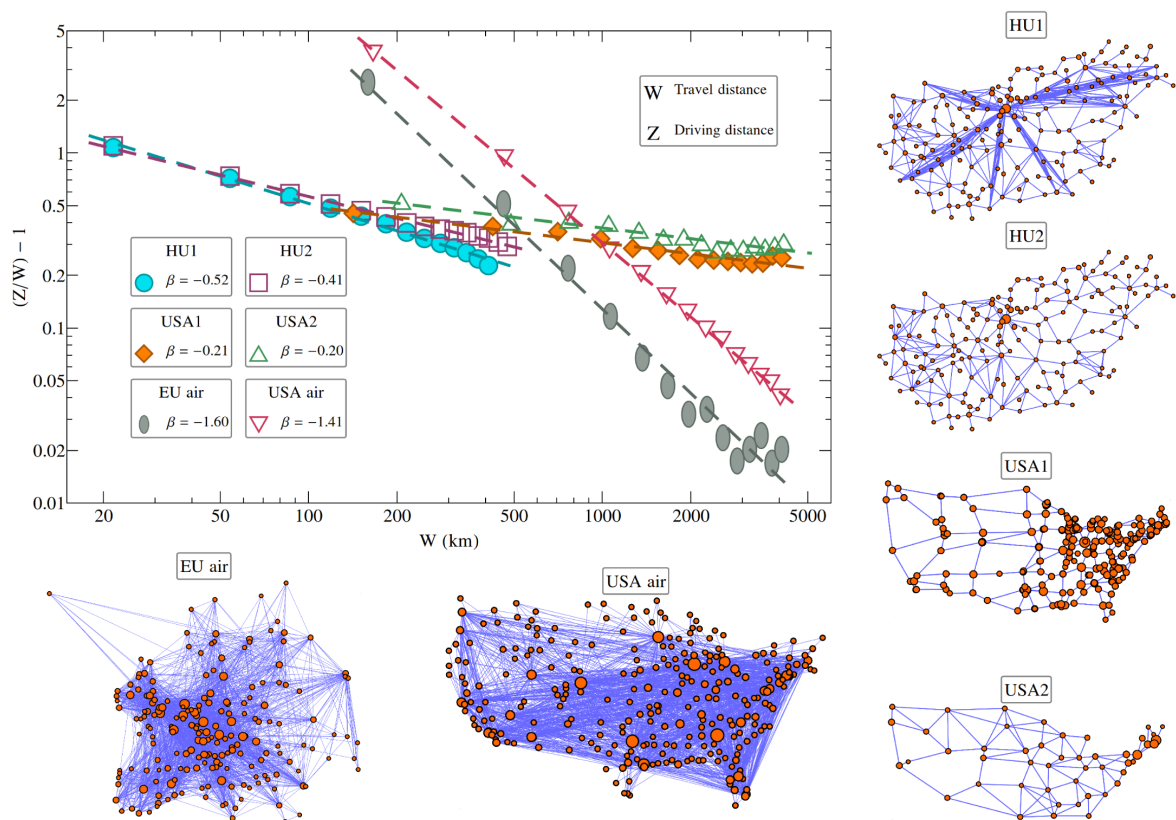


Figure 2.2: Topologia rețelelor de călătorie. Scalarea între distanța parcursă (z) și distanța geodezică (w) pe diferite rețele de călătorie. Liniile sugerează validitatea relației (2.1). De asemenea, figura ilustrează structura topologică a unor rețele de transport uman utilizate în studiul de față. De remarcat axele logaritmice.

călătoriile pe drumurile naționale și autostrăzi, precum și la zboruri directe între aeroporturi. Rezultatele pentru călătoriile cu mașina au fost calculate pe rutele naționale și autostrăzile din Ungaria, respectiv rutele interstatale din SUA continentală. Distribuția spațială a locațiilor și structura rețelelor rutiere este prezentată în Figura 2.2. În cazul Ungariei, am luat în considerare călătoriile între orașele a 174 subregiuni statistice pe teritoriul țării [29]. Aici arătăm rezultate distincte pentru cazurile cu și fără traseu permis pe autostrăzi (hărți HU1 și HU2, respectiv). În cazul SUA am studiat 241 locații în vecinătatea punctelor de joncțiune ale principalelor drumuri interstatale (harta SUA1) și 48 capitalele de stat (harta USA2).

Pentru călătoriile cu avionul, am considerat doar zboruri directe din aeroporturile din Europa, SUA și întreaga lume. Poziția spațială a acestor aeroporturi este ilustrată pe hărțile din Figura 2.2. Pentru Europa, am folosit datele de timp și coordonatele GPS ale 203 aeroporturi majore [30], după cum se vede pe harta aeriană a UE. În cazul SUA, am lucrat cu 282 aeroporturi de pe teritoriul SUA (harta aeriană a SUA), dar numai cu datele topologice ale rețelei de călătorii aeriene, deoarece nu aveam acces la datele de zbor. Prin urmare, am căutat și alte date, cu urmărire pe GPS, înregistrate de 500 de zboruri din întreaga lume [31]. Aceste date nu sunt afișate pe harta spațială și trebuie să remarcăm că datele aeriene și hărțile din SUA sunt utilizate pentru calcularea distanței, iar datele GPS aeriene au fost utilizate pentru calculul vitezei de deplasare.

În consecință, cele două panouri inserate din Figura 2.1 afișează datele privind călătoria cu mașina și cu

avionul pentru viteza aparentă de deplasare în funcție de distanța de deplasare. Am obținut o tendință de tip lege de putere cu exponentul de scalare de 0.07 pentru drumuri și 0.25 pentru zboruri, indicat prin linii punctate. Cu toate acestea, există o diferență vizibilă în cazul diferitelor infrastructuri rutiere, mai precis în cazul datelor HU1 este o tendință de creștere mai ridicată a vitezei aparente, datorită utilizării atât a drumurilor cât și a autostrăzilor (pentru harta rutieră a se vedea Figura 2.2).

Datorită topologiei specifice a rețelelor rutiere, de obicei nu există linii drepte între orașe, deci trebuie să distingem două tipuri de distanțe. Distanța geodezică (w) înseamnă distanța pe liniile geodezice între punctul sursă și țintă, iar distanța parcursă, notată cu z se referă la lungimea minimă a drumului din rețea. Având în vedere diferitele distanțe, putem defini viteze diferite. Viteza aparentă, notată cu v , este calculată ca fiind (distanța geodezică)/(timpul de călătorie), în timp ce viteza de deplasare (u) este calculată ca (distanța parcursă)/(timpul călătoriei).

Să analizăm mai întâi relația dintre aceste două distanțe. În timpul analizei topologice a rețelelor de transport schițate în Figura 2.2, constatăm că distanța geodezică și cea parcursă, în medie, pot fi exprimate (una în funcție de cealaltă) sub forma unei relații de scalare

$$\frac{z}{w} - 1 = C \cdot w^{-\beta} \quad (2.1)$$

unde exponentul β al funcției de putere exprimă relația dintre distanța parcursă și distanța geodezică.

După cum este ilustrat în Figura 2.2, exponenții obținuți $\beta \approx 1.4 - 1.6$ (pentru călătoria cu avionul) și $\beta \approx 0.2 - 0.5$ (pentru călătoria cu mașina) sugerează că, pentru călătoriile cu avionul, valoarea z converge mai rapid la w decât în cazul călătoriilor rutiere, adică, căile de zbor sunt mai apropiate de liniile geodezice decât drumurile.

Tendința de creștere a vitezei de deplasare în funcție de distanța de rulare rezultă din efectul combinat al celor două tipuri de întârziere. În primul rând, în ambele capete ale unei traiectorii, nodurile sursă și țintă, generează întârzieri. În cazul nostru, nodurile sunt orașe în care complexitatea traficului determină întârzieri mai mici și mai mari, în funcție de mărimea orașelor. În al doilea rând, întârzieri se produc și pe parcurs.

Rezultatele noastre susțin ipoteza potrivit căreia cu cât mai departe călătorim, cu atât mai repede mergem. Toate datele de mai sus demonstrează creșterea vitezei aparente medii în funcție de distanță atât pentru diferite moduri de deplasare, cât și luate împreună.

În a **doua** tematică legată de mobilitatea umană oferim o generalizare a modelului de radiație, luând în considerare și efectul costurilor de călătorie [12], și examinăm aplicabilitatea acestui model pe o bază de date construită din datele navetiștilor din Ungaria.

Modelul original de Radiație (RM) [10] se bazează pe presupunerea simplă că persoanele care caută un loc de muncă își optimizează veniturile acceptând cea mai apropiată ofertă dintre locurile de muncă care oferă un salariu mai bun decât cel disponibil în prezent în locația lor actuală. Presupunând o funcție de distribuție cumulativă $p_{\leq}(z)$ pentru veniturile din societatea studiată, probabilitatea $P_{>}(z|n)$ ca o persoană cu venit z să refuze cele mai apropiate n oferte de locuri de muncă este

$$P_{>}(z|n) = [p_{\leq}(z)]^n \quad (2.2)$$

Folosind funcția de densitate a probabilității pentru venituri $p(z) = \frac{\partial p_{\leq}(z)}{\partial z}$, putem calcula acum probabilitatea de a nu accepta cele mai apropiate n joburi. Apoi, acceptând ipoteza că numărul de locuri de muncă pe un teritoriu este proporțional cu populația W ($n = \mu W$), în cadrul modelului de radiație obținem

$$P_{>}(W)_{RM} = \frac{1}{\mu W + 1} \text{ (un parametru)} \quad (2.3)$$

pentru ca o persoană să se deplaseze într-o locație, care este în afara unui disc centrat pe locația sa actuală, și care conține o populație totală W .

Presupunând că persoanele aflate în căutarea unui loc de muncă sunt selective în alegerea lor și că sunt dispuși să accepte oferte de muncă mai bune doar cu o probabilitate q , obținem modelul de Radiație cu Selecție (RMwS) [11]. În urma calculelor modelului de radiație original cu criteriile de selecție nou introduse, obținem (pentru mai multe detalii, consultați teza):

$$P_{>}(W)_{RMwS} = \frac{1 - (1 - q)^{\mu W + 1}}{(\mu W + 1) q} \text{ (model cu doi parametri)} \quad (2.4)$$

Pentru $q = 1$, regăsim modelul original de radiație. Modelul RMwS descrie mai bine distribuția fluxurilor de navetiști din SUA cu cei doi parametri ai săi [11]. Aceasta nu este totuși o surpriză, deoarece ne-am aștepta, în mod firesc, ca un model cu doi parametri să ofere o regresie mai bună decât un model cu un parametru. Pornind din nou de la modelul original de radiație, am introdus o altă variantă a acestuia, pe care l-am numit modelul de radiație optimizat pentru costurile de călătorie (TCORM). Spre deosebire de modelul original de radiație, unde acceptarea postului este independentă de distanța dintre reședință și locul de muncă, aici introducem o condiție mai puternică pentru navetă: indivizii vor alege naveta dacă există un venit mai bun după scăderea costurilor de călătorie. În acest fel, costurile de deplasare depind și de distanța parcursă și nu numai de numărul n a joburilor care este pe traiectoria de tranzit. Folosind această ipoteză și urmând raționamentul ipotezelor modelelor de radiații obținem (pentru mai multe detalii, consultați teza):

$$P_{>}(W)_{TCORM} = \frac{1 + \lambda\sqrt{W}}{\mu W + 1} \text{ (model tot cu doi parametri)} \quad (2.5)$$

Pentru testarea modelelor prezentate mai sus, procesăm o bază de date completă a navetiștilor și a populației din Ungaria. Am analizat datele recensământului populației pe anul 2011 [32] și, de asemenea, am utilizat date cu o rezoluție de 1 km^2 pentru distribuția populației în anul recensământului 2011 [33].

În timpul procesării datelor, selectăm unul câte unul așezările i ca sursă pentru navetă și construim discurile cu raza $d(i, j)$, ajungând la destinația j . Aceasta este schițată în Figura 2.3. Numărăm populația totală $w_i[j]$ în interiorul acestui disc și înregistrăm numărul de navetiști $f_i(j)$ care pleacă de la așezarea i și călătoresc către așezarea j .

Cu datele $d(i, j)$, $f_i(j)$ și $w_i[j]$ pentru toate perechile de așezări (i, j) se calculează probabilitățile experimentale $P_{>}(W)$.

Folosind algoritmi de analiză a datelor bazate pe notațiile de mai sus, trecem la construirea experimen-

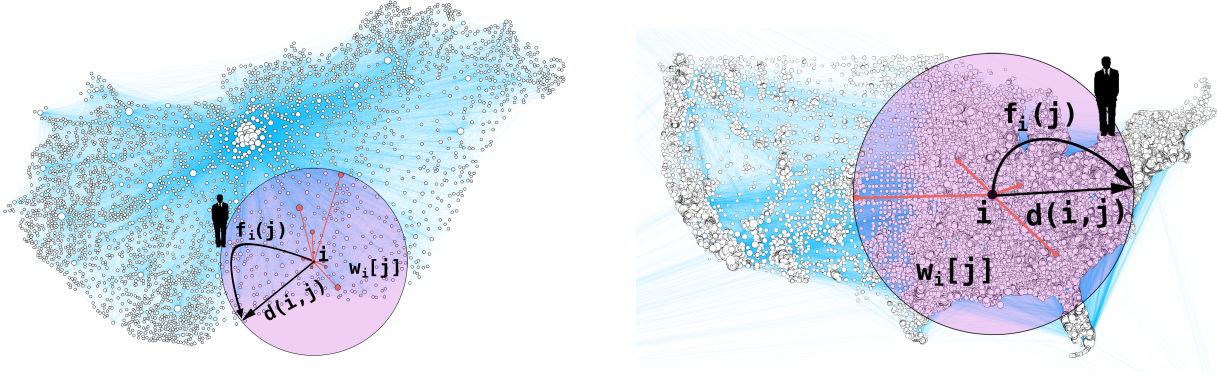


Figure 2.3: Notatii și metode de procesare a datelor în rețeaua de navetă din Ungaria (stânga) și SUA (dreapta). Sunt construite discuri de diferite raze $d(i, j)$, pornind de la o anumită așezare și ajungând la celelalte așezări j . Populația $w_i[j]$ în interiorul acestor discuri și numărul navetiștilor care pleacă din locația i și care călătoresc către locația j , $f_i(j)$ este înregistrată.

tală a curbei $P_{>}(W)$ din datele recensământului din 2011. Valorile calculate sunt comparate cu rezultate obținute din modelul original RM (2.3), modelul RMwS (2.4) și modelul nostru TCORM (2.5). Efectele de element finit (prezența hotarelor) devin importante pentru valori mari ale W (discurile centrate pe așezări devin în mare parte incomplete, deoarece se extind peste granițele Ungariei). În consecință, am considerat datele numai până la $W = 1\,000\,000$.

Rezultatele obținute indică faptul că modelul TCORM funcționează mai bine decât modelele simple RM și RMwS. Modelele RM și RMwS sunt capabile doar să descrie o parte din intervalul relevant pentru W , spre deosebire de modelul TCORM, care oferă o ajustare vizuală bună pentru întregul interval.

În **ultima** tematică investigată în cadrul mobilităților umane, introducem un alt model pentru fluxurile de navetiști și extindem investigația folosind densități de populație la scară largă și fluxuri de navetiști și din alte regiuni geografice [13]. Pe baza analizelor prezentate mai sus, prelucrăm baze de date complete din SUA, Italia și Ungaria.

Pe lângă modelul RM (2.3), modelul RMwS (2.4) și modelul TCORM (2.5), studiem și comportamentul modelului de Gravitație (GM) [9]. De asemenea, introducem și testăm modelul nostru original de tip “flow and jump” (FJM).

Modelul GM presupune că numărul de navetiști $f_i(j)$ între orașele i și j se poate aproxima prin:

$$f_i(j) = F(W_i) \frac{(W_j)^\alpha}{(r_{i,j})^\beta} \quad (2.6)$$

unde W_i este populația așezării i și $r_{i,j}$ este distanța dintre așezările i și j . $F(x)$ este o funcție de kernel ce crește monoton, iar α și β sunt exponenți obținuți din fitare.

Folosind datele $f_i(j)$ putem calcula și probabilitatea $P_{>}^i(W)_{GM}$, ca o persoană care trăiește în locația i să se deplaseze într-o locație care se află în afara unui disc centrat pe locația sa actuală și care conține o populație W :

$$P_{>}^i(W)_{GM} = 1 - \frac{\sum_{j \neq i}^{(w_i[j] < W)} f_i(j)}{\sum_j f_i(j)} = 1 - \frac{\sum_{j \neq i}^{(w_i[j] < W)} \frac{(W_j)^\alpha}{(r_{i,j})^\beta}}{\sum_j \frac{(W_j)^\alpha}{(r_{i,j})^\beta}} \quad (2.7)$$

Această probabilitate este independentă de funcția kernel $F(x)$. Am notat cu $w_i[j]$ populația totală din interiorul discului. Acum, probabilitatea medie ca navetiștii să călătorească în afara discului cu o populație de W este

$$P_{>}(W)_{GM} = \langle P_{>}^i(W)_{GM} \rangle_i \quad (2.8)$$

Introducem acum un nou model doar cu un parametru ca alternativă pentru modelul RM simplu. Îl denumim modelul de “flow and jump” (FJM). Modelul se bazează pe o simplă ecuație master pentru densitatea de probabilitate $\rho(n, t) = -dP_{>}(n, t)/dn$. Folosim aici aceleași notații ca și cele pentru modelul RM. Pe baza ipotezelor modelelor de tip “creștere și resetare”, care sunt introduse în articolul de sinteză [34], presupunem un proces invers: un flux de probabilitate înspre înapoi, suplimentat de un proces de salt de la origine la o stare oarecare având valoarea n . Ecuația master continuă este următoarea (pentru mai multe detalii, consultați [34]):

$$\frac{d\rho(n, t)}{dt} = \frac{\partial(\eta(n)\rho(n, t))}{\partial n} + [\gamma(n)\rho(n, t)]\rho(0, t) \quad (2.9)$$

Această ecuație master (2.9), specifică un proces în care există un debit local de densitate de probabilitate netă din fiecare stare către starea $n = 0$ și o probabilitate de salt de la origine ($n = 0$) la o stare n . Pentru ratele $\eta(n)$ (rata de flux) și $\gamma(n)$ (rata de salt) dependente de stare considerăm kerneluri simple care sunt realiste pentru procesul de navetă. Tranzițiile $0 \rightarrow n$ guvernate de ratele $\gamma(n)\rho(n, t)$ descriu probabilitatea ca muncitorii să aleagă un loc de muncă care presupune navetă. $\gamma(n)$ ar trebui să descrească cu distanța, iar proporționalitatea cu $\rho(n, t)$ sugerează că locurile populare de navetă au multe locuri de muncă bune. Prin urmare, alegem următoarele forme pentru $\eta(n)$ și $\gamma(n)$:

$$\rho_s(n) = \frac{\eta(0)\rho_s(0)}{\eta(n)} e^{-\int_0^n \frac{\gamma(x)\rho_s(0)}{\eta(x)} dx} \quad (2.10)$$

După efectuarea calculelor, pe baza presupunerii $n(r) = \mu W(r)$ suntem conduși la o formulă care seamănă cu probabilitatea obținută în cadrul modelului de radiație (pentru mai multe detalii consultați teza):

$$P_{>}(W)_{FJM} = \frac{1}{(\mu W + 1)^{(a-1)}} \quad (2.11)$$

În cele ce urmează, considerăm modelul FJM cu parametrul universal $a = 7/4$, care oferă o fitare mult îmbunătățită pentru datele reale de navetă. În cazul $a = 2$, regăsim modelul original de radiație. Modelul este unul cu doi parametri, totuși, dacă setăm valoare $a = 7/4$, devine similar cu modelul RM, un model cu un singur parametru.

Pentru a verifica modelul în afara datelor disponibile pentru Ungaria, folosim un set de date pe scară mai largă pentru SUA și un set de date de dimensiuni mai mici pentru Italia.

Pentru SUA am analizat datele estimate ale recensământului populației între 2006 și 2010 [35] folosind $Q = 73\,803$ așezări (noduri) (cercurile albe din Figura 2.3) și 4 156 426 rute de navetă (legături) (linii albastre între cercurile albe din Figura 2.3). Pentru studierea distribuției spațiale a populației, am folosit o bază de date pentru anii 2006 și 2010. Această bază de date a furnizat populația estimată a SUA continen-

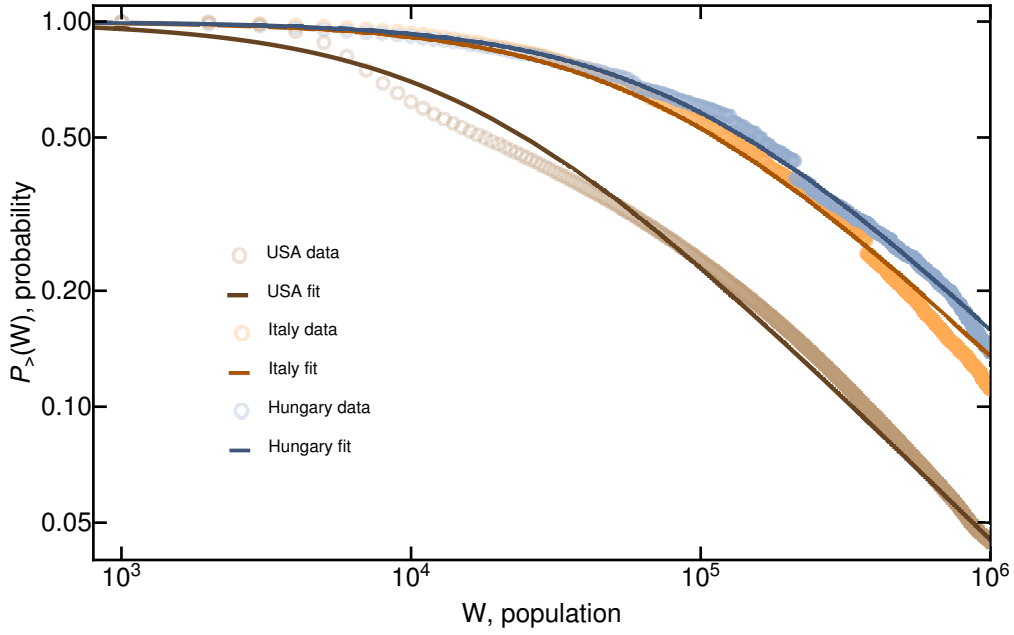


Figure 2.4: Comparație vizuală între predicția modelului FJM și datele experimentale pentru toate cele trei țări (SUA, Italia, Ungaria). Liniile, compuse din cercuri, prezintă datele experimentale $P_{>}(W)$ și liniile colorate întunecate sunt cele fitate pentru predicția modelului FJM (2.11). Pentru $a = 7/4$ valorile fitate pentru μ sunt date în Tabelul 2.1.

	USA	Italy	Hungary
μ	0.000062	0.000013	0.000011
R^2	0.993	0.997	0.998

Table 2.1: Parametrii de fitare și performanța fitărilor arătate în Figura 2.4, având în vedere forma funcțională dată de ecuație (2.11) și $a = 7/4$.

tală împărțită în 11 078 286 de celule de suprafețe de 1 km^2 [36]. Pentru a accelera calculele, am “renormalizat” spațial și am obținut o rezoluție mai puțin exactă cu celule de dimensiune de 4 km^2 . Acest lucru se realizează prin colapsarea datelor a patru celule vecine și prin medierea coordonatelor latitudinale și longitudinale. Ca rezultat, am obținut 1 230 920 de celule, care conțin o populație totală $W = 308\,745\,231$.

Datele din Italia conțin $Q = 8\,093$ așezări, 556 120 rute de navetă și provin din recensământul populației italiene realizat în 2011 [37]. Populația totală $W = 55\,605\,065$ este distribuită în celule de 1 km^2 suprafață [38].

Procesarea datelor experimentale se bazează pe pașii schițați deja pentru Ungaria. Cu datele $d(i, j)$, $f_i(j)$ și $w_i[j]$ pentru toate perechile de așezări (i, j) (vedeți detalii pe Figura 2.3) calculăm probabilitățile $P_{>}(W)$.

În primul pas, probabilitățile $P_{>}(W)$ calculate pentru SUA sunt comparate cu cele mai bune rezultate obținute din modelul original RM (2.3), modelul RMwS (2.4), modelul TCORM (2.5), modelul GM (2.8) și noul model FJM (2.11). În modelul FJM, parametrul $a = 7/4$ este fixat pentru toate seturile de date studiate, astfel încât singurul parametru liber al acestui model este μ . Pentru modelul GM fitarea a fost realizată luând în considerare o metodă de tip “mash” pentru diferite valori $\alpha \in [-1.0, 2.5]$ și $\beta \in [-1.0, 2.5]$.

Pentru a minimiza efectul de mărime finită (influența marginilor), examinăm datele doar până la $W_{max} = 1\,000\,000$, iar pentru rutele scurte de navetă stabilim un prag mai mic de $W_{min} = 1\,000$. Fitarea este efectuată în intervalul $[W_{min}, W_{max}]$.

Statisticile obținute sunt în favoarea modelelor FJM și GM. Modelul FJM oferă o descriere bună a datelor experimentale studiate. Faptul că modelul FJM este mai performant decât modelul RM, se datorează parametrului fixat la $a = 7/4$. Cu toate acestea, modelul FJM depășește și modelele cu doi parametri, RMwS și TCORM. De asemenea, modelul GM studiat oferă o fitare bună, dar în timpul fitării pentru diferite țări nu putem fixa parametrii α și β .

Modelul FJM pentru $a = 7/4$ funcționează mai bine și pentru datele de navetă prelucrate pentru Ungaria și Italia. Calitatea fitării și parametrii fitării sunt prezentate în Tabelul 2.1. În plus, Figura 2.4 arată fitarea FJM pentru curbele $P_{>}(W)$ determinate experimental pentru toate cele trei baze de date investigate.

Rezultatele noastre sugerează că dacă se utilizează modele de tip RM, este important să se țină seama de faptul că selectarea locurilor de muncă este dependentă de distanță, cost și dimensiune.

Rezultatele fitării prin modelul FJM pentru SUA, Italia și Ungaria sunt însumate în Tabelul 2.1, unde cel mai optim parametru μ caracterizează atât disponibilitatea locurilor de muncă pentru populație, cât și atractivitatea acestor locuri de muncă pentru solicitanți.

“Cu cât este mai complexă rețeaua, cu cât tiparul de interconexiuni este mai complex, cu atât mai rezistent va fi.”

Fritjof Capra

3

Dinamica citărilor și rețele complexe

Acest capitol este dedicat studiului rețelilor științifice de publicații și a rețelilor sociale. Ca un prim pas prezentăm metodele noastre de detectare a comunităților bazată pe diagrame grafice Voronoi și pe partiționare stohastica de tip Voronoi [39].

Pe baza metodologiei universal acceptate, dacă algoritmul de detectare a comunității este pregătit pentru teste, îl considerăm mai întâi pe diferite benchmark-uri și după acela pe rețele din lumea reală. Benchmark-urile sunt generate cu ajutorul software-ului de referință furnizat de Lancichinetti și colab. [40, 41]. Acest prim pas este necesar, deoarece de aici știm exact care dintre noduri aparțin fiecărui cluster. În schimb, în cazul rețelilor din lumea reală nu cunoaștem, de obicei, afilierea nodurilor la cluster, dar este important ca algoritmi noștri să funcționeze bine și în rețelele din lumea reală.

Ca o problemă specifică în care detectarea de comunități este importantă, am studiat normalizarea indicatorilor scientometrici în cazul publicațiilor individuale. Am propus aici un algoritm local de detectare a clusterelor pentru a identifica comunitatea scientometrică a unui articol. După ce se detectează clusterul local, se calculează mai mulți indicatori ai articolelor pe aceasta comunitate. Se aplică apoi o metodă de normalizare pentru aceste valori.

Tot în cadrul acestei tematici am studiat statistica și dinamica evoluției legăturilor pe publicațiile științifice și pe rețeaua de prietenie Facebook sugerând existența a două legi simple: legarea preferențială și creșterea exponențială a numărului de noduri în rețea. Am obținut că distribuția share-urilor pentru posturile de pe Facebook și distribuția citărilor științifice se pot fi bine cu ajutorul funcției de distribuție Tsallis-Pareto cu exponentul $g = 1.4$ [20].

Prezentăm succint ca un **prim** rezultat o soluție geometrică pentru a detecta comunități pe grafuri, bazate pe diagrame Voronoi [39, 42]. Această metodă este folosită, de obicei, pentru a partiționa spații metrice în regiuni (celule Voronoi) în jurul punctelor selectate (puncte generatoare), așa cum este ilustrat în Figura 3.1A. Fiecare punct al spațiului aparține punctului generator cel mai apropiat. Aplicăm aceasta soluție în cazul grafurilor, unde toate legăturile au o lungime pozitivă și distanța dintre două noduri este egală cu cea mai scurtă rută dintre ele, vezi Figura 3.1B.

Pentru a introduce metoda de clusterizare, mai întâi prezentăm metoda partiționării Voronoi în spațiul Euclidian 2D. Considerăm un set de puncte într-un plan 2D, distribuit astfel încât să formeze grupuri locale. Calculăm densitatea locală a punctelor din interiorul teritoriilor și selectăm punctele generatoare ale celulelor Voronoi din interiorul teritoriului lor cu raza r , așa cum se vede în Figura 3.1A. Fiecare punct

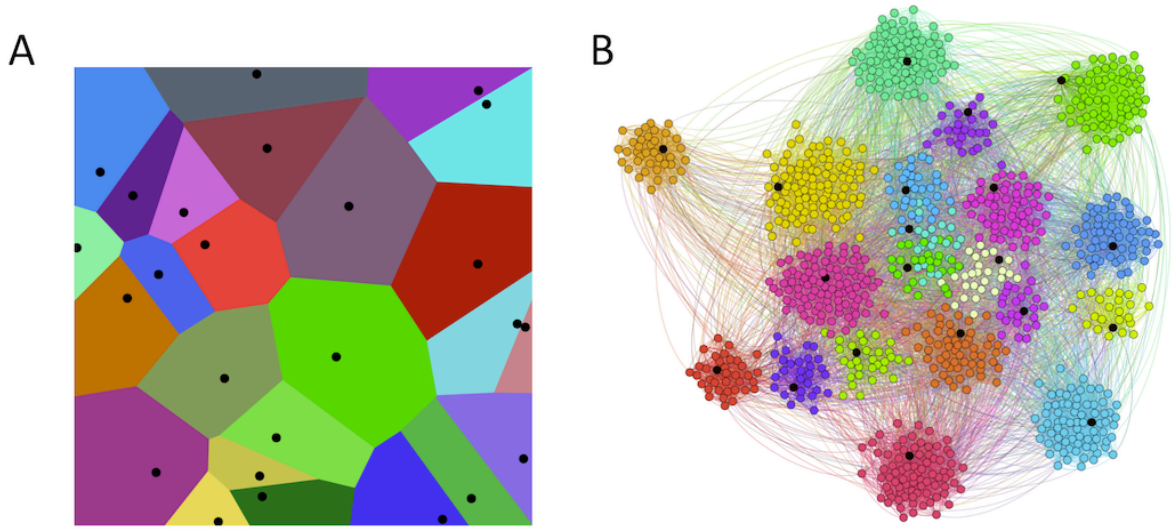


Figure 3.1: Diagrame Voronoi. (A) Ilustrația partiționării Voronoi în spațiul Euclidian 2D. (B) Diagrame Voronoi pentru grafuri reprezentat de aplicația grafică Gephi folosind algoritmul schemă ForceAtlas2 [43]. Nodurile generatoare sunt afișate în negru.

din plan este atribuit la o celulă Voronoi cu punctul generator cel mai apropiat. Ca rezultat, punctele sunt împărțite în grupuri de către celulele Voronoi.

Pentru a aplica teoria anterioară pentru rețele, trebuie să transformăm graful în spațiu metric. Mai întâi calculăm o valoare a distanței între fiecare două noduri, care este lungimea celei mai scurte rute dintre acestea. Lungimea oricărei rute este egală cu suma lungimii legăturilor de-a lungul rutei, unde am definit lungimea de legătură ca fiind inversul coeficientului de grupare a legăturilor (ECC) prezentat în [44]. ECC-ul unui legături între nodul i și nodul j este calculat ca

$$C_{i,j} = \frac{z_{i,j} + 1}{\min[(k_i - 1), (k_j - 1)]} \quad (3.1)$$

unde k_i, k_j sunt numerele de legături a nodurilor, $z_{i,j}$ este numărul de triunghiuri de care aparține legătura și $\min[\dots]$ este numărul de triunghiuri potențiale la care ar putea să aparțină, deoarece acesta este valoarea mai mică decât numerele de legături a celor două noduri adiacente, minus unu (legătura examinată). Dacă valoarea $1/C_{i,j}$ este mare, atunci legătura probabil conectează noduri din diferite clustere.

După ce am definit distanța, următorul pas este să determinăm punctele generatoare. În timpul calculului am folosit o metodă de selecție a punctelor generatoare Voronoi bazată pe densitatea locală relativă a nodurilor [45]. Aceasta funcționează pe un subgraf format din primii vecini ai nodului i . Pentru acest subgraf determinăm

$$\rho_i = \frac{m}{m + k} \quad (3.2)$$

unde m este numărul de legături din interiorul vecinătății și k este numărul de legături care ies din subgraf.

Dacă am determinat punctele generatoare în funcție de distanța $1/C_{i,j}$ și raza r , atunci am obținut un graf clusterizat. Este evident că variația parametrului r va influența numărul de comunități din graf. Am studiat influența valorii r și am obținut că o valoare relativ mică pentru r asigură o bună partiționare

în ambele cazuri: pentru benchmark-uri și în rețelele din lumea reală. Cea mai bună strategie pentru partiționarea Voronoi este totuși creșterea valorii r și examinarea calității clusterizării, de exemplu prin modularitatea clusterului [46].

Am testat metodele noastre pe benchmark-uri și rețele din lumea reală. Am folosit algoritmi de generare a benchmark-urilor cu o mare varietate de setări ai diferiților parametri [40, 41]. Am generat 100 de grafuri cu diferiți parametri. Aceste grafuri au fost partiționate prin creșterea valorilor r și s-au monitorizat funcțiile de calitate pentru partiționarea noastră și pentru comunitățile predefinite pe grafic. Pentru comparație am folosit funcția de modularitate [46] și rezultatul informației reciproce dintre clusterizarile noastre.

Când metoda a atins valoarea maximă a modularității, valorile obținute din informațiile reciproce au fost apropiate de 1 (pentru valoarea optimă a lui r), indicând faptul că metoda a identificat cu succes comunitățile originale pe benchmark-uri. Cu toate acestea, se observă că în unele cazuri structura comunităților nu este foarte clară, datorită faptului că există o mulțime de grupări mici.

Contrar cazului benchmark-urilor, în cazul rețelilor din lumea reală gruparea este mai complicată. În acest caz nu cunoaștem structurile optime ale comunităților, iar succesul metodei noastre poate fi judecată doar prin compararea cu alte metode deja acceptate. Am testat algoritmul nostru pe mai multe rețele din lumea reală, cu structură și origine diferite, și rezultatele le-am comparat cu rezultatele altor cinci algoritmi folosiți pe scară largă. Am obținut că metoda noastră funcționează bine pentru valori relativ mici ai parametrului r , și numai în cazul în care avem o pronunțată amestecare între comunități, nu găsim clusterelor potrivite.

Ca și a **doua** problematică, considerăm o versiune stohastică a metodei de detectare a comunității de mai sus [47]. Esența noii metode este selectarea aleatorie a punctelor generatoare și folosirea matricei de coeziune Voronoi. Fiecare nod are posibilitatea de a fi un punct generator. Pentru o anumită selecție a nodurilor generatoare, se calculează matricea de coeziune Voronoi, care este probabilitatea de co-localizare a unei perechi de noduri, indicând probabilitatea perechilor intra-comunitare și inter-comunitare. În forma considerată, valorile sunt mai mari pentru perechile intra-comunitare decât pentru perechile inter-comunitare. În urma acestei metode se obține o clusterizare prin niște pași stochastici.

Știind că fiecare nod poate fi selectat aleator ca un punct generator, detaliem acum pașii algoritmului:

- alegem în mod aleatoriu un număr de g noduri generatoare și efectuăm o partiționare graf-Voronoi pe distanța de-a lungul celei mai scurte rute. Apoi, pentru fiecare pereche de noduri, determinăm co-localia acestora.
- repetăm această partiționare de R_e ori și calculăm matricea de coeziune Voronoi, care este definită ca fiind media matricelor de co-localie. Aceste matrici de coeziune sunt reprezentate astfel încât nodurile să fie ordonate conform informațiilor “ground truth”.
- am identificat faptul că nodurile intra-modulare și inter-modulare sunt oarecum separate. Pornind de la aceasta, aplicăm o modificare iterativă a topologiei rețelei. Un procent mic al legăturilor cu coeziune scăzută este mutat între nodurile neconectate cu coeziune asociată ridicată. Ca rezultat al modificării, structura comunității este păstrată, însă separarea dintre ele este definită mai clar.

Am testat algoritmul nostru pe benchmark-uri. Am generat un benchmark cu mai multe niveluri, cu

28 și respectiv 10 comunități pe primul și pe al doilea nivel. Am evaluat matricea de coeziune și relocarea topologică la fiecare ciclu. Ca rezultat, am obținut 9 comunități în ambele cazuri (cu și fără relocare topologică), dar matricea de coeziune indică o mai bună clusterizare prin ordonarea topologică.

Înainte de a testa rețelele din lumea reală, să vedem cum se construiește matricea de coeziune Voronoi. Un element al matricei este o probabilitate pentru co-localizarea unei perechi de noduri, adică probabilitatea ca cele două noduri să aparțină aceeași celule Voronoi. Matricea conține probabilități diagonale și non-diagonale. Urmând metoda ordonării după informațiile “ground truth”, nodurile aparținând aceluiași cluster sunt situate cu probabilitate mai mare pe diagonală matricei.

Dacă este necesar, atunci când matricea de coeziune nu prezintă clar comunitățile, folosim o metodă de mărire a contrastului, în decursul căreia mutăm topologic legăturile slabe, și după aceea separăm comunitățile printr-o valoare de prag. În timpul metodei de mărire a contrastului, comunitățile pot fi determinate după cum urmează: toate nodurile primesc o etichetă de comunitate separată, iar într-o buclă, peste toate nodurile, eticheta comunitară a nodului curent este atribuită acelor noduri care nu și-au schimbat încă eticheta și a căror coeziune cu nodul curent depășește un anumit prag.

Clusterizarea rețelor din lumea reală rămâne dificilă. În timpul detectării comunității, metoda stohastică simplă graf-Voronoi generează harta de coeziune, unde diferența dintre inter și intra perechi nu este semnificativă.

În lumina acestora, metoda stohastică simplă graf-Voronoi funcționează bine pe benchmark-uri. În cazul rețelor din lumea reală este necesară și tehnica de mărire a contrastului. În unele cazuri extreme, se impune și metoda combinată.

În **continuarea** acestor studii, studiem normalizarea indicatorilor scientometrici ai publicațiilor individuale [19]. Indicatorii scientometrici populari, cum ar fi factorul de impact [48], eigenfactor [49] sau h-index [50] nu pot asigura o comparație directă a diferitelor discipline. De asemenea, în cadrul aceluiași jurnal, numărul de citări pentru articolele publicate sunt foarte diferite.

Reprezentarea publicațiilor și citărilor prin intermediul grafurilor de publicații este larg utilizată. Fiecare publicație și citare în această rețea este indicată de noduri și de legături între ele. Folosind această identificare, indicatorii scientometrici individuali devin calculabili în astfel de rețele. Mai precis, evaluarea articolului este realizată prin numărul citărilor, adică prin numărul legăturilor care intră în nod sau prin versiunea locală de PageRank [51] interpretat strâns legat de domeniile științifice detectate.

Pe baza ideii metodei paralelizate de detectare a comunităților locale [52], introducem un algoritm de detectare locală a clusterelor (LCD). Aceasta este o metodă de expansiune în care plecând de la o publicație, algoritmul detectează comunitățile fără a partiționa întreaga rețea. Am testat algoritmul LCD atât pe benchmark-uri, cât și pe rețele din lumea reală.

După finalizarea detectării comunităților calculăm indicatorii scientometrici relevanți pentru comunitatea locală detectată. Primul indicator este numărul de citări a unui articol, ceea ce reflectă impactul său asupra comunității științifice. Acest indicator este egal cu gradul de intrare al nodurilor, adică cu numărul de referințe din întreaga rețea. Al doilea indicator este versiunea simplificată (locală) de PageRank [51].

Prima cantitate statistică importantă care trebuie normalizată este statistica citărilor articolelor studiate. În cazul nostru, distribuția probabilității unui indicator de articol reprezintă probabilitatea de apariție a

articolelor prin valorile indicatorului de articol. Prin urmare, aceste distribuții de probabilități vor descrie comportamentul citărilor în diferite domenii științifice.

Menționăm că gradul de intrare n_i al clusterelor pe aceeași benchmark-uri are o distribuție care este o dreaptă pe figura log-log

$$p(n_i) = An_i^\alpha \quad (3.3)$$

unde A este o constantă de proporționalitate. Pentru simplificare, redimensionăm valorile n_i , astfel încât fiecare distribuție să poată fi fitată cu constanta de normalizare $A = 1$. În consecință, fiecare grad de intrare trebuie să fie înmulțit cu un factor de scalare $\xi = A^{1/\alpha}$, după cum urmează:

$$p(n_i) = An_i^\alpha = (A^{1/\alpha} n_i)^\alpha = (\xi n_i)^\alpha \quad (3.4)$$

Această metodă de redimensionare este testată pe diferite rețele de citări din lumea reală, de exemplu în rețeaua de citare Web of Science cu 771 914 articole și 7 779 703 citări. Am pornit algoritmul LCD, apoi am măsurat distribuțiile pentru gradul de intrare și pentru PageRank-ul local. Distribuțiile gradelor de intrare arată că în cazul rețelei WoS este necesar să se scaleze numărul de citări pentru a obține o distribuție normalizată cu constanta de proporționalitate $A = 1$.

Funcția de distribuție locală PageRank, studiată în mod similar, este de asemenea fitată cu o funcție de putere. Datele sunt fitate cu funcția $p(n_i) = 10^{-7} n_i^{-1.4}$. Scalarea și factorii de scalare obținuți diferă semnificativ datorită naturii diferite a indicatorului de articol PageRank local.

Ca un **ultim** studiu descris în acest capitol, prezentăm pe scurt rezultatele noastre privind statisticile citărilor și a share-urilor pe Facebook [53]. Este bine cunoscut faptul că citările evaluează articolele, autorii sau instituțiile. Urmând modelul citărilor științifice, de asemenea, share-urile pe Facebook evaluează posturile sau utilizatorii. Citările, publicațiilor sau distribuirea posturilor cuantifică și caracterizează calitatea acestora. În cercetarea noastră ne concentrăm asupra acestor distribuții și căutăm legi de universalitate.

Pornind de la studiile anterioare care dezvăluie unele universalități în distribuția citărilor [54], am analizat și noi distribuția citărilor și a share-urilor. Studiile anterioare au arătat că citările venite de la mai multe instituții academice și reviste se rescalează pe o curbă comună dacă ele sunt normalizate în raport cu valoarea medie. Cu alte cuvinte, s-a calculat densitatea de probabilitate $f(x)$ pentru un articol cu x citări și s-a reprezentat valoarea $\langle x \rangle f(x)$ în funcție de $x/\langle x \rangle$. $\langle x \rangle$ este valoarea medie a lui x . Rezultatele prezentate în Figura 3.2 arată că diferitele seturi culese după diferite metodologii se rescalează pe aceeași curbă.

De asemenea, observăm că distribuțiile prezintă o tendință clară de funcție de putere pentru valorile mari ale numărului de citări, în domeniul $x/\langle x \rangle > 10$. Am arătat că întreaga distribuție poate fi fitată cu succes cu distribuția de tip Tsallis-Pareto (TP) [20]

$$f(x) = \frac{g}{(g-1)\langle x \rangle} \left(1 + \frac{x}{(g-1)\langle x \rangle} \right)^{-1-g} \quad (3.5)$$

care este o funcție de densitate de probabilitate (PDF) cu o coadă de funcție de putere. Aceasta nu este o distribuție riguros independentă de scală, această proprietate este satisfăcută numai pentru valori $g > 1$

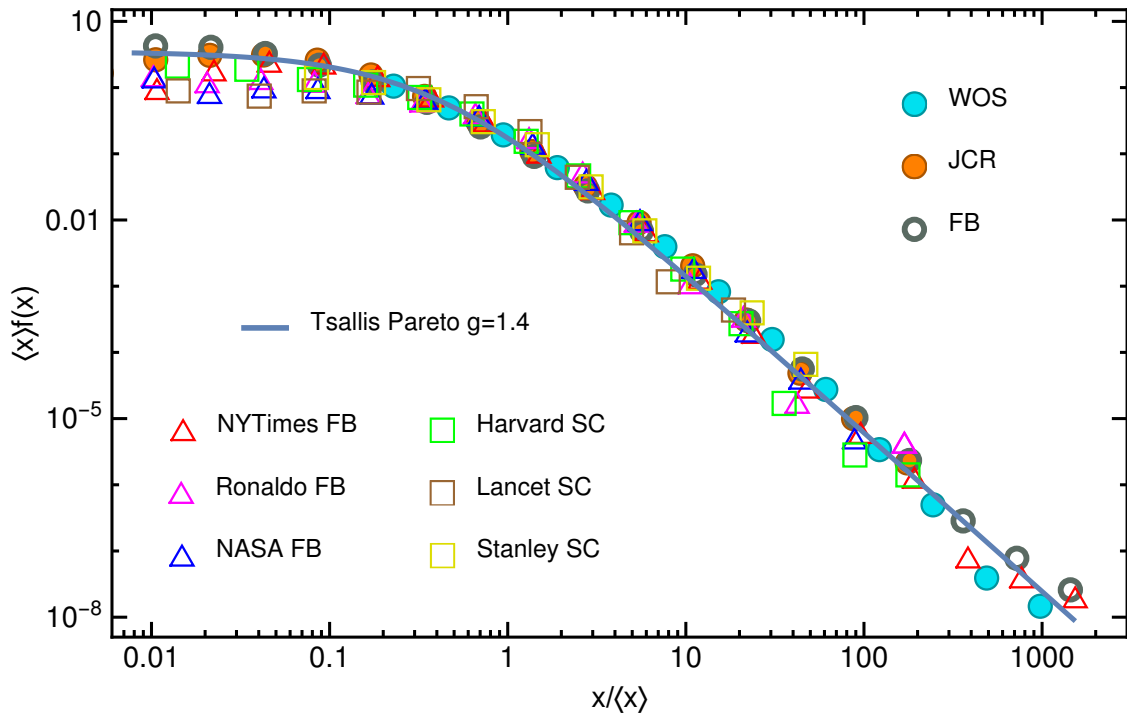


Figure 3.2: Distribuția a numerelor de citare (share-uri). $f(x)$ este densitatea de probabilitate (PDF) pentru o lucrare (post) care are x citate/share-uri. Prezentăm valoarea $\langle x \rangle f(x)$ în funcție de $x/\langle x \rangle$ ($\langle x \rangle$ valoarea medie sau primul moment al PDF). Pentru un număr ridicat de citări, este vizibilă o tendință clară de scalare. Simboluri diferite sunt pentru diferite seturi de date, așa cum este ilustrat în legendă. Întreaga curbă poate fi bine fitată cu o distribuție TP (3.5) cu $g \approx 1.4$ și $\langle x \rangle = 1$.

și valori $x/\langle x \rangle$ suficient de mari. Prin urmare, este mai adecvată denumirea acestor distribuții ca fiind distribuții cu coadă grea (heavy tail) [55].

Presupunând validitatea acestei distribuții universale, am procesat mai multe seturi de date, așa cum se arată în Figura 3.2. În primul rând, am procesat mai mult de 600 000 publicații ISI Web of Science (WoS) [56], statisticile citărilor pe zece ani pentru toate citările indexate ISI (aprox. 12 000) din Journal Citation Report (JCR) [57] și mai mult de 150 000 postări de la 16 utilizatori de Facebook (pagini) diferiți [58]. Toate cele trei seturi de date sugerează că funcția de densitate de probabilitate de tipul TP cu un singur parametru este adecvat pentru fitarea datelor cu parametrul $g \approx 1.4$, așa cum se arată prin linia continuă și punctele de date din Figura 3.2.

Pe lângă seturile de date mari, am colectat date de pe 16 pagini Facebook. Prezentăm trei tipuri diferite de pagini, unul de știri, New York Times (NYTimes FB), unul a unei celebrități sportive, Cristiano Ronaldo (Ronaldo FB), și unul de știință, pagina NASA (NASA FB). În cazul fiecăreia, este vizibil faptul că funcția de densitate de probabilitate de tip TP fitează foarte bine și aceste seturi mai mici de date.

În plus, am studiat citările articolelor publicate în 1990, având autori de la Universitatea Harvard (Harvard SC), citările articolelor publicate în revista The Lancet (Elsevier) în 1990 (Lancet SC), precum și citările pentru un singur autor din domeniul fizicii, Prof. H. E. Stanley de la Universitatea din Boston, SUA, care are 965 publicații ISI și 62 996 citări ISI (Stanley SC). Și aceste date se fitează bine prin PDF-ul de tip TP.

Pentru toate seturile de date, am construit funcția de distribuție de probabilitate experimentală printr-o metodă logaritmică de grupare pe clase cu binuri de dimensiune 2^n . În afară de fluctuațiile observabile în cazul seturilor mici de date, toate datele colectate au urmat aceeași tendință cu $g \approx 1.4$.

Pentru a modela rezultatele obținute, am elaborat un model dinamic bazat pe o ecuație de tip masters. S-au folosit două presupuneri, care au fost experimental verificate: creșterea exponențială a numărului de citări (sau share-uri) și creșterea preferențială (efectul Matei). Pentru cazul Facebook, creșterea exponențială este evidențiată de prezentarea lui Mark Zuckerberg, potrivit căreia și activitatea de distribuire a informațiilor pe Facebook crește exponențial [59]. Efectul Matei – “celui ce are, i se va da” – a fost discutat în numeroase lucrări anterioare pentru mai multe sisteme sociale, inclusiv cel al citărilor științifice [60, 61].

4

Concluzii

În cadrul tezei am studiat dinamica în sistemele sociale folosind metode specifice ale fizicii computaționale. Sarcina dificilă cu care ne-am confruntat a fost sintetizarea logică și prezentarea într-o manieră pedagogică a studiilor pe care le-am efectuat în domeniile conexe: jocuri evolutive cu participanți multipli, mobilități umane, rețele sociale și scientometrie. Pe lângă modelare, o provocare cu care ne-am confruntat a fost extragerea de date, prelucrarea datelor și dezvoltarea de instrumente numerice și analitice pentru compararea modelului cu datele din lumea reală.

Am studiat mai multe tipuri de jocuri evolutive care utilizează reguli de ajustare a strategiilor de imitare și logit pentru diferite tipuri de rețele. Am investigat distribuțiile strategiilor diferitelor tipuri de jucători prin simulări Monte Carlo și calcule analitice.

Am studiat corelații în distribuțiile strategiilor, am aplicat reguli de sincronizate a ajustării strategiilor și am demonstrat procesele de invazie și de speciație. S-a ilustrat impactul pozitiv al regulilor de ajustare prin imitare și logit pentru menținerea cooperării. Apoi, am examinat invazia pe trei direcții: orizontale, verticale și înclinate.

Ca un al doilea subiect, am analizat legile universale care guvernează mobilitățile umane și am modelat fluxurile de navetiști în funcție de distanță. Am studiat diferite mobilități umane pe rețele rutiere și pe rețele de transport aerian. Am investigat tiparurile de navetă între localități în trei arii geografice și am elaborat două modele originale. Pentru a compara într-o manieră critică modelele noastre teoretice cu realitatea, am folosit rețele de transport uman, seturi de date de recensământ, informații asupra locurilor de muncă și asupra distanțelor.

Ca rezultat important, am confirmat ipoteza noastră conform căreia există o regulă universală: “cu cât călătorim mai departe, cu atât mai rapid mergem”. Pe baza distanței de deplasare și a timpului de călătorie necesar, am demonstrat că viteza medie aparentă crește în funcție de distanță, urmând o tendință asemănătoare unei legi de putere.

Apoi am studiat distribuția fluxurilor de navetiști. Am folosit modele de tip radiație, modelul de gravitație și modelul “flow and jump” (flux și salt) pentru fitarea datelor experimentale observate. Modelul “flow and jump” propus de noi a oferit cele mai bune rezultate pentru toate cele trei seturi de date experimentale.

În cele din urmă, am investigat reprezentarea comunităților și efectele de clusterizare ale rețelilor. Am

propus o normalizare bazată pe domenii pentru articolele științifice și am arătat reguli similare de scalare pentru publicațiile științifice și pentru posturile de pe Facebook. Studiul a fost realizat pe rețele de publicații și respectiv sociale. Am testat algoritmi noștri pe benchmark-uri și pe rețelele din lumea reală.

Pentru rețelele științifice bazate pe publicații am propus două metode de detectare a comunităților. Aceste metode pot fi folosite pentru gruparea eficientă a nodurilor în diferite clustere. Am dezvoltat o metodă locală de detectare a clusterelor, care funcționează fără a procesa întreaga rețea. Acest lucru s-a dovedit a fi important pentru determinarea și normalizarea indicatorilor individuali. Pentru a realiza acest lucru, am normalizat distribuțiile de probabilitate ale indicatorilor științifici în concordanță cu proprietățile datelor de a scală pentru a elimina diferențele dintre domeniile științifice.

În afară de aceste studii, s-a urmărit și comportamentul referitor la citări în rețelele de publicații și la distribuirea informației pe rețelele sociale. Am observat că ele prezintă același tip de tipar de popularitate. În urma unei abordări printr-o ecuație master, am modelat statisticile caracteristice în ambele sisteme și am constatat că funcția de distribuție a probabilității de tip Tsallis-Pareto descrie bine distribuțiile normalizate ale acestora.

În concluzie, considerăm că teza introduce abordări noi pentru a caracteriza dinamica sistemelor sociale, dezvăluind aspecte noi și interesante ale acestor sisteme, unele necunoscute până în prezent.

Publicații legate de teză

5.1 ARTICOLE ȘTIINȚIFICE

- Szabó, Gy., **Varga, L.**, and Borsos, I. Evolutionary matching-pennies game on bipartite regular networks. *Phys. Rev. E*, 89, 042820 (2014).
- Deritei, D., Lázár, Zs. I., Papp, I., Járai-Szabó, F., Sumi, R., **Varga, L.**, Ravasz Regan, E., and Ercsey-Ravasz, M. Community detection by graph Voronoi diagrams. *New J. Phys.*, 16, 1–17 (2014).
- **Varga, L.**, Vukov, J., and Szabó, Gy. Self-organizing patterns in an evolutionary rock-paper-scissors game for stochastic synchronized strategy updates. *Phys. Rev. E*, 90, 042920 (2014).
- Vukov, J., **Varga, L.**, Allen, B., Nowak, M. A., and Szabó, Gy. Payoff components and their effects in a spatial three-strategy evolutionary social dilemma. *Phys. Rev. E*, 92, 012813 (2015).
- **Varga, L.**, Kovács, A., Tóth, G., Papp, I., and Néda, Z. Further We Travel the Faster We Go. *PLoS ONE*, 11, e0148913 (2016).
- Szabó, Gy., **Varga, L.**, and Szabó, M. Anisotropic invasion and its consequences in two-strategy evolutionary games on a square lattice. *Phys. Rev. E*, 94, 052314 (2016).
- **Varga, L.**, Tóth, G., and Néda, Z. An improved radiation model and its applicability for understanding commuting patterns in Hungary. *Regional Statistics*, 6, 27–38 (2016).
- Lázár, Zs. I., Papp, I., **Varga, L.**, Járai-Szabó, F., Deritei, D., and Ercsey-Ravasz, M. Stochastic graph Voronoi tessellation reveals community structure. *Phys. Rev. E*, 95, 022306 (2017).
- Néda, Z., **Varga, L.**, and Biró, T. S. Science and Facebook: The same popularity law! *PLoS ONE*, 12, e0179656 (2017).
- **Varga, L.**, Tóth, G., and Néda, Z. Commuting patterns: The Flow and Jump model and supporting data. *EPJ Data Science*, 7, 37 (2018).

5.2 ARTICOLE APĂRUTE ÎN PROCEEDINGS-URI ALE UNOR CONFERINȚE

- **Varga, L.**, Deritei, D., Ercsey-Ravasz, M., Florian, R., Lázár, Zs. I., Papp, I., and Járai-Szabó, F. Normalizing scientometric indicators of individual publications using local cluster detection methods on citation networks. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 12(9), 1189–1198 (2018).

5.3 PREZENTĂRI LA CONFERINȚE

- “Commuting patterns: a modified radiation model and supporting data”, MaCS 2018, 12th Joint Conference on Mathematics and Computer Science, Săcuieu, Romania, 14-17 June 2018.
- “Normalizing scientometric indicators of individual publications using local cluster detection methods on citation networks”, ICCSIB 2018 : 20th International Conference on Cybermetrics, Scientometrics, Informetrics and Bibliometrics, Barcelona, Spain, 29-30 October 2018.

5.4 POSTERE PREZENTATE LA CONFERINȚE

- “Local Cluster Detection Method for Normalizing Scientometric Indicators” and “Community detection using stochastic graph Voronoi tessellation”, NetSci 2015, Zaragoza, Spain, 1-5 June 2015.
- “Testing human mobility models with commuters data” and “Velocity versus distance scaling in human travel”, XXXVI Dynamic Days Europe, Corfu, Greece, 6-10 June 2016.
- “Genetic-like algorithm applied on citation networks for evaluating scientific publications”, Conference of Complex Systems 2016, Beurs Van Berlage, Amsterdam, The Netherlands, 19-22 September 2016.
- “An improved radiation model and its applicability for understanding commuting patterns”, meco43: 43rd Conference of the Middle European Cooperation in Statistical Physics, Kraków, Poland, 1-4 May 2018.

Bibliografie selectată

- [1] T. Parsons. *The Social System*. Routledge, London, 1991.
- [2] M. Barthélemy. Spatial networks. *Phys. Rep.*, 449:1–101, 2011.
- [3] J. Maynard Smith. *Evolution and the Theory of Games*. Cambridge University Press, Cambridge, UK, 1982.
- [4] C. Domb. Ising model. In C. Domb and M. S. Green, editors, *Phase Transitions and Critical Phenomena, Vol. 3*, pages 357–484. Academic Press, London, 1974.
- [5] Gy. Szabó and G. Fáth. Evolutionary games on graphs. *Phys. Rep.*, 446:97–216, 2007.
- [6] K. Binder and D. W. Heermann. *Monte Carlo Simulation in Statistical Physics. An Introduction*. Springer-Verlag, Berlin, 1992.
- [7] R. Gallotti and M. Barthélemy. Anatomy and efficiency of urban multimodal mobility. *Sci. Rep.*, 4:6911, 2014.
- [8] M. C. González, C. A. Hidalgo, and Barabási A. L. Understanding individual human mobility patterns. *Nature*, 453:779–782, 2008.
- [9] F. Lukermann and P. W. Porter. Gravity and potential models in economic geography. *Contributions to probability and statistics*, 50:493–504, 1960.
- [10] F. Simini, M. C. González, A. Maritan, and A. L. Barabási. A universal model for mobility and migration patterns. *Nature*, 484:96–100, 2012.
- [11] F. Simini, A. Maritan, and Z. Néda. Human mobility in a continuum approach. *PLoS ONE*, 8(3):e60069, 2013.
- [12] L. Varga, G. Tóth, and Z. Néda. An improved radiation model and its applicability for understanding commuting patterns in hungary. *Regional Statistics*, 6:27–38, 2016.
- [13] L. Varga, G. Tóth, and Z. Néda. Commuting patterns: The flow and jump model and supporting data. *EPJ Data Science*, 7:37, 2018.
- [14] A. L. Barabási. *Network Science*. Cambridge University Press, United Kingdom, 2012.
- [15] M. E. J. Newman. *Networks: An Introduction*. Oxford University Press, New York, 2010.
- [16] R. N. Kostoff. Citation analysis cross-field normalization: a new paradigm. *Scientometrics*, 39(3): 225–230, 1997.
- [17] F. Radicchi, S. Fortunato, and C. Castellano. Universality of citation distributions: Toward an objective measure of scientific impact. *Proc. Natl Acad. Sci.*, 105(45):17268–17272, 2008.
- [18] A. Schubert and T. Braun. Cross-field normalization of scientometric indicators. *Scientometrics*, 36(3):311–324, 1966.

- [19] L. Varga, D. Deritei, M. Ercsey-Ravasz, R. V. Florian, Zs. I. Lázár, I. Papp, and F. Járai-Szabó. Normalizing scientometric indicators of individual publications using local cluster detection methods on citation networks. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 12(9):1189–1198, 2018.
- [20] S. Thurner, F. Kyriakopoulos, and C. Tsallis. Unified model for network dynamics exhibiting nonextensive statistics. *Phys. Rev. E*, 76:036111, 2007.
- [21] L. E. Blume. The statistical mechanics of strategic interactions. *Games Econ. Behav.*, 5:387–424, 1993.
- [22] L. E. Blume. The statistical-mechanics of best-response strategy revision. *Games Econ. Behav.*, 11: 111–145, 1995.
- [23] Gy. Szabó, L. Varga, and I. Borsos. Evolutionary matching-pennies game on bipartite regular networks. *Phys. Rev. E*, 89:042820, 2014.
- [24] R. J. Glauber. Time-dependent statistics of the Ising model. *J. Math. Phys.*, 4:294–307, 1963.
- [25] L. Varga, J. Vukov, and Gy. Szabó. Self-organizing patterns in an evolutionary rock-paper-scissors game for stochastic synchronized strategy updates. *Phys. Rev. E*, 90:042920, 2014.
- [26] J. Vukov, L. Varga, B. Allen, M. A. Nowak, and Gy. Szabó. Payoff components and their effects in a spatial three-strategy evolutionary social dilemma. *Phys. Rev. E*, 92:012813, 2015.
- [27] Gy. Szabó, L. Varga, and M. Szabó. Anisotropic invasion and its consequences in two-strategy evolutionary games on a square lattice. *Phys. Rev. E*, 94:052314, 2016.
- [28] L. Varga, A. Kovács, G. Tóth, I. Papp, and Z. Néda. Further we travel the faster we go. *PLoS ONE*, 11:e0148913, 2016.
- [29] Hungarian central statistical office, regional atlas - Hungary's territorial breakdown by statistical micro-regions for regional development. https://www.ksh.hu/regional_atlas_microregions, 2015.
- [30] T. Dusek. The time frame of the air travel in Europe. *Private communication*, 2009.
- [31] GPSlib, GPS tracks hosting service. <http://www.gpslib.net>, 2015.
- [32] Census tract flow, commuting data, Hungary. <http://www.ksh.hu>, 2011.
- [33] Population distribution, Hungary. <http://ec.europa.eu/eurostat/cache/GISCO/geodatafiles/GEOSTAT-grid-POP-1K-2011-V2-0-1.zip>, 2011.
- [34] T. S. Biró and Z. Néda. Unidirectional random growth with resetting. *Physica A*, 499:355–361, 2018.
- [35] CTPP Census tract flows, commuting data, American Community Survey. https://www.fhwa.dot.gov/planning/census_issues/ctpp/data_products/2006-2010_tract_flows/, 2006-2010.
- [36] Population distribution, American Community Survey. <https://www.census.gov/geo/maps-data/data/tiger-data.html>, 2006-2010.
- [37] Census tract flow, commuting data, Italy. http://www.istat.it/storage/cartografia/matrici_pendolarismo/matrici_pendolarismo_2011.zip, 2011.

- [38] Population distribution, Italy. <http://ec.europa.eu/eurostat/cache/GISCO/geodatafiles/GEOSTAT-grid-POP-1K-2011-V2-0-1.zip>, 2011.
- [39] F. Aurenhammer. Voronoi diagrams – a survey of a fundamental geometric data structure. *ACM Comput. Surv.*, 23:345–405, 1991.
- [40] A. Lancichinetti and S. Fortunato. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys. Rev. E*, 80:016118, 2009.
- [41] A. Lancichinetti, S. Fortunato, and F. Radicchi. Benchmarks graphs for testing community detection algorithms. *Phys. Rev. E*, 78:046110, 2008.
- [42] D. Deritei, Zs. I. Lázár, I. Papp, F. Járai-Szabó, R. Sumi, L. Varga, E. Ravasz Regan, and M. Ercsey-Ravasz. Community detection by graph voronoi diagrams. *New J. Phys.*, 16:1–17, 2014.
- [43] M. Bastian, S. Heymann, and M. Jacomy. Gephi: an open source software for exploring and manipulating networks. *Proceedings of ICWSM: AAAI Int. Conf. on Weblogs and Social Media*, pages 361–362, 2009.
- [44] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi. Defining and identifying communities in networks. *Proc. Natl Acad. Sci.*, 101:2658–2663, 2004.
- [45] S. Fortunato. Community detection in graphs. *Phys. Rep.*, 486:75–174, 2010.
- [46] M. E. J. Newman. Modularity and community structure in networks. *Proc. Natl Acad. Sci.*, 103:8577–8582, 2006.
- [47] Zs. I. Lázár, I. Papp, L. Varga, F. Járai-Szabó, D. Deritei, and M. Ercsey-Ravasz. Stochastic graph voronoi tessellation reveals community structure. *Phys. Rev. E*, 95:022306, 2017.
- [48] E. Garfield. The impact factor and using it correctly. *Der Unfallchirurg*, 101(6):413–414, 1998.
- [49] C. T. Bergstrom, J. D. West, and M. A. Wiseman. The eigenfactor metrics. *Journal of Neuroscience*, 28(45):11433–11434, 2008.
- [50] J. E. Hirsch. An index to quantify an individual’s scientific research output. *Proc. Natl Acad. Sci.*, 102(46):16569–16572, 2005.
- [51] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the Web. *Technical Report, Stanford InfoLab*, 1999-66, 1999.
- [52] J. P. Bagrow and E. M. Bollt. Local method for detecting communities. *Phys. Rev. E*, 72(4):046108, 2005.
- [53] Z. Néda, L. Varga, and T. S. Biró. Science and facebook: The same popularity law! *PLoS ONE*, 12:e0179656, 2017.
- [54] A. Chatterjee, A. Ghosh, and B. K. Chakrabarti. Universality of citation distributions for academic institutions and journals. *PloS ONE*, 11:0146763, 2016.
- [55] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51:661–703, 2009.
- [56] Web of Science. <http://www.webofknowledge.com>, 2015.
- [57] InCites, Journal citation report, Thomson Reuter. <https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action>, 2016.

- [58] Facebook. <https://www.facebook.com>, 2016.
- [59] M. Zuckerberg. Online sharing is growing at an exponential rate. <https://www.youtube.com/watch?v=HNy9uxcRedU>, 2012.
- [60] H. Jeong, Z. Nédá, and A. L. Barabási. Measuring preferential attachment in evolving networks. *Europhys. Lett.*, 61:567–572, 2003.
- [61] J. Wang. Unpacking the Matthew effect in citations. *J. of Informetrics*, 8:329–339, 2014.