

Analiza Computațională a Comportamentelor Emergente în Rețele

Teză de doctorat – Rezumat

pentru obținerea titlului științific de doctor la

Universitatea Politehnica Timișoara

în domeniul de doctorat Calculatoare și Tehnologia Informației

autor: ing. Gabriel BARINA

conducător științific: prof.univ.dr.ing. Mircea VLĂDUȚIU

Domeniul interdisciplinar Network Science se ocupă cu analiza rețelelor sociale, facilitând o mai bună înțelegere a societății noastre, atât din punct de vedere fizic, cât și social. Astfel, odată cu studiul comportamental al agenților (sociali, muzicali, economici, *etc.*), putem obține o mai bună înțelegere a proceselor de interacțiune, colaborare și influențare.

Prezenta teză se bazează pe cercetarea proceselor menționate cu ajutorul sistemelor de calcul. Astfel, folosind calculatorul ca instrument de bază în analiza rețelelor, propun simularea modelelor complexe emergente într-o manieră dinamică, pentru o mai bună înțelegere a dinamicității sociale înconjurătoare, a comportamentului uman și a modului în care acestea se influențează reciproc.

Obiectivele principale ale acestei teze sunt:

1. Crearea unei rețele colaborative emergente folosind puterea analitică a calculatorului (*e.g.* data-mining, machine-learning, *etc.*) în vederea analizei și compararea acesteia cu alte rețele similare.
2. Propunerea unei metrici capabile de cuantificare a sociabilității unor agenți sau rețele.
3. Crearea unui simulator capabil de a modela și simula relațiile și interacțiunile emergente dintre agenții economici.
4. Folosirea simulatorului pentru analiza venitului cumulativ al agenților economici, a distribuției veniturilor și a ergodicității rețelelor economice.

Pentru îndeplinirea primelor două obiective, analizez în prima parte a tezei, rețeaua emergentă formată din muzicieni. Prin aplicarea unor metode tradiționale de analiză ale rețelelor complexe, respectiv a elementelor autentice expuse odată cu teza de față, analizez rețeaua muzicienilor din punct de vedere colaborativ (*i.e.* dintre muzicieni), dar și din punct de vedere economic (*i.e.* activitatea lor de a produce conținut muzical). Ca atare, elementele de originalitate aduse cu aceasta teză sunt următoarele:

- Obținerea datelor folosind Big Data mining: datorită naturii datelor necesare pentru acest studiu, informațiile au fost colectate din mai multe baze de date online. Adicional, inspirat din studiul muzicienilor Jazz, MuSeNet nu este limitat la un singur gen muzical, ci ia în considerare muzicieni din toate genurile muzicale.
- Analiza centralităților: am analizat rețeaua MuSeNet din perspectiva mai multor metrici de centralități importante prin metode computaționale.
- Segregarea și suprapunerea comunităților existente: folosind puterea analitică a calculatorului, am segregat comunitățile fundamentale folosind un algoritm de învățare automată (machine learning) dedicat detecției de comunități. Adicional, am identificat suprapunerea genurilor, am detectat agenții influenți și am determinat “Kevin Bacon”-ul industriei muzicale.

- Analiza motif-ilor: am aplicat o metodă originală de analiză și diferențierea rețelelor, identificând motif-urile existente, pentru a extrage proprietățile topologice ale rețelelor sociale.
- *S*-metric: generând algoritmic o metrică nouă pentru diferențierea rețelelor similare din punctul de vedere al sociabilității lor, am determinat sociabilitatea rețelei MuSeNet, cât și a altor rețele.

În cea de-a doua parte a tezei prezint un simulator socio-economic original, bazat atât pe observații empirice, cât și pe modele economice inovatoare. Având capacități precise de simulare, este folosit la analiza proceselor de colaborare și interacțiune, dar și la distribuția emergentă de venit la scară macro, cu reguli specifice bine-definite în prealabil la scară micro. Ca atare, contribuțiile aduse în a doua parte a tezei țin de simulatorul original socio-economic, cu următoarele proprietăți:

- Algoritm euristic personalizat: cu îmbunătățiri aduse la mecanismul de modelare comportamentală a agenților economici, inspirându-mă din modelul de interacțiune bazat pe toleranță.
- Implementarea teoriilor economice importante: adăugând realism și flexibilitate pentru modelarea agenților, a deciziilor și a interacțiunilor.
- Caracteristici aplicabile în lumea reală: folosirea oricărei topologii pentru simulări, crearea și evoluția dinamică a agenților, a rolurilor, a produselor, a prețurilor, investițiilor, *etc.*
- Simulator de seamă: folosind TrEcSim, am analizat următoarele attribute ale sistemelor economice integrate:
 - Distribuțiile statice și dinamice de venit: am analizat comportamentele dinamice și emergente dintre agenți.
 - Ergodicitatea: am determinat corectitudinea intrinsecă a sistemelor economice coroborată cu conceptul de ergodicitate din fizica statistică, bazat atât pe topologia rețelelor, cât și pe plasarea producătorilor în interiorul rețelei.

Ca și structură, această teză include un capitol introductiv, unul cu fundamente elementare teoretice, și un ultim capitol în care descriu state-of-the-art-ul domeniului ales. În următoarele două capitole, descriu cele două rețele colaborative emergente analizate (MuSeNet și TrEcSim), fiecare având un pasaj dedicat. Teza se încheie cu un capitol dedicat concluziilor, contribuțiilor, respectiv direcției de cercetare, bibliografie (219 titluri consultate și citate) și o anexă. Această teză se întinde pe 122 pagini și conține 62 figuri, respectiv 15 tabele.

În primul capitol prezint o scurtă introducere în domeniul Network Science. Conținând elemente din științe exacte – *e.g.* știința calculatoarelor, fizică, matematică, *etc.* – Network Science facilitează studiul societății noastre, a comportamentelor și relațiilor prin folosirea calculatorului ca unealtă pentru modelarea și simularea datelor de tip Big Data. În mod similar, și Social Network Analysis a stârnit interesul comunității științifice, ramura principală a Network Science, datorită aplicabilității acesteia în analiza rețelelor reale sau sintetice, atât din punct de vedere topologic (*i.e.* felul în care nodurile se interconectează), cât și comportamental (*i.e.* felul în care aceste noduri interacționează).

Motivația mea în alegerea acestui domeniu multi-disciplinar este una întemeiată: încă din perioada masteratului am fost expus la acest domeniu relativ tânăr și la posibilitatea de a folosi tehnici și elemente însușite din știința calculatoarelor pentru analiza rețelelor complexe, respectiv identificarea proprietăților dinamice, structurale, colaborative sau emergente. În acest context, teza de față contribuie la elucidarea fenomenelor și mecanismelor sociale emergente

(e.g. influențare, colaborare, formarea unei noi entități, *etc.*), cât și la identificarea proprietăților dinamice sau structurale. Într-adevăr, domeniul interdisciplinar Network Science contribuie la o înțelegere îmbunătățită a comportamentelor sociale și economice, demonstrând importanța înțelegerii acestora. nu doar în diferite științe sociale, dar și în inginerie și tehnologie.

În cel de-al doilea capitol prezint câteva elemente teoretice necesare pentru înțelegerea lucrării de față. Astfel, prezint o clasificare a rețelelor complexe în patru mari categorii (*i.e.* biologice, sociale, tehnologice și semantice), a topologiilor de bază (*i.e.* regular mesh, random, small-world și scale-free) și a metricilor specifice rețelelor complexe (*i.e.* centrality, degree distribution, average path length, clustering coefficient, modularity), folosite în această teză. În mod conex cu aceste metrici, prezint de asemenea două abordări de mare interes în analiza rețelelor, și anume: așa-numitele network motif-uri și metric fidelity.

O proprietate comună a tuturor rețelelor este posibilitatea reprezentării și analizării lor cu sub-grafuri, denumite oficial motif. Recurente și importante din punct de vedere statistic pentru rețeaua din care fac parte, acestea pot fi folosite și la compararea diferitor rețele. Cu toate acestea, detecția lor prezintă o oarecare dificultate; ca atare, în studiul de față mă folosesc doar de cele formate din trei noduri.

O altă metodă populară de a analiza și a compara în mod cantitativ rețele - metoda folosită și în această teză -, este de a determina fidelitatea metrică a rețelelor și compararea lor. Creată pentru a putea exprima similitudinea dintre rețele, poate fi aplicată și pentru determinarea similitudinii rețelelor sintetice cu cele din lumea reală.

În capitolul al treilea prezint noutățile din domeniul studiat, referitoare la procesul de colaborare. În termeni generali, rețelele complexe sunt formate dintr-un set de actori (noduri) interconectați în baza unor reguli specifice. Aceste noduri, deși autonome, distribuite din punct de vedere geografic, și eterogene din punct de vedere al mediului de operare, obiective culturale, sociale, *etc.*, prezintă aceeași proprietate de bază: dorința de a colabora pentru a obține rezultate comune. Astfel, colaborarea este cu atât mai preferată, cu cât beneficiile finale ce pot fi obținute sunt mai semnificative. Din punctul de vedere al structurii rețelelor, procesul de colaborare poate fi reprezentat printr-o legătură fizică între nodurile participante în acest proces. În acest context, acest capitol este dedicat prezentării studiilor relevante (din diferite domenii), ce tratează procesul de colaborare cu ajutorul rețelelor complexe:

- **Rețeaua autorilor de publicații științifice:** un reper contemporan în cercetarea academică a procesului de colaborare, reprezintă studiul autorilor diferitor publicații științifice. Acești autori nu mai sunt agenți izolați, ci fac parte dintr-o rețea colaborativă multi-disciplinară. Printre altele, analiza acestor rețele ne oferă răspuns referitor la numărul articolelor publicate, calitatea lor, locația autorilor, dar și evoluția rețelei în timp. Ca structură, această rețea este de tip small-world, cu comunități clar delimitate. Acest lucru se datorează faptului că acțiunea de colaborare este adoptată de la un nod (autor) la altul, în funcție de aria științifică a grupei respective.
- **Universul Marvel:** universul Marvel, în cei peste 70 de ani de existență, s-a expandat într-un ritm atât de alert, încât s-a folosit de puterea grafurilor pentru a crea un overview complet cu toți super-eroii, relațiile dintre aceștia, cronologia evenimentelor, *etc.* Astfel, s-a creat o legătură între super-eroi care au apărut, chiar și pentru scurt timp, în același film, desene, serial, joc, *etc.* Rețeaua obținută – formată din ~6,500 noduri (super-eroi) și ~10,000 muchii (relații) –, deși una sintetică, se aseamănă foarte mult cu alte rețele formate în mod natural, diferind doar clustering coefficient-ul, datorită distribuției super-eroilor. Acest lucru vine în contradictoriu cu rețeaua autorilor, datorită originii artificiale a rețelei.
- **Rețeaua actorilor IMDB:** Similar cu studiul precedent, s-a procedat și cu rețeaua

actorilor IMDB, creând legătura între ei în funcție de colaborarea lor în filme sau seriale. Conex, comunitatea științifică și-a concentrat atenția și asupra afirmației faimoase a lui Kevin Bacon, cum că el ar fi colaborat - în mod direct, ori indirect - cu toți actorii de la Hollywood. Ca rezultat, s-a introdus o nouă metrică în literatură, și anume numărul Bacon; cu cât acest număr este mai mare, cu atât respectivul actor este mai departe de Kevin Bacon în rețea.

- **Rețeaua muzicienilor Jazz:** Într-o manieră similară s-a procedat și cu studiul ce analizează procesul de colaborare dintre muzicienii jazz. Astfel, autorii acestui studiu au creat legături atât între muzicieni, cât și între trupe, dacă au cel puțin un membru comun. Unul dintre cele mai interesante rezultate obținute, reprezenta segregarea muzicienilor în comunități, fie din cauza discriminării rasiale, fie datorită amplasării lor geografice.

În al patrulea capitolul analizăm MuSeNet, rețeaua colaborativă a muzicienilor. Astfel, pornind de la premisele prezentate în capitolul anterior, cât și din dorința de a mă folosi de rețele complexe pentru a observa și analiza relațiile dintre muzicieni, am creat rețeaua Musenet. Datele necesare au fost colectate din mai multe surse, obținând informații referitoare la ~20,000 muzicieni și ~5,000 trupe.

Network Analysis

În Figura 4.2 putem observa comunitățile relevante emergente din MuSeNet în funcție de genul muzical, și anume Pop/Rock 24,56%, Jazz 16,72%, Blues 15,8%, Classical 8% și Country 5,35%. Deși aceste proporții nu reprezintă neapărat o noutate, elementele de noutate în sine constau în distribuția spațială, cât și în suprapunerea acestor genuri muzicale. Ca atare, genurile populare sunt atât centrale, cât și mai grupate, datorită colaborării dintre muzicieni. Pe de altă parte, cu cât genurile muzicale sunt mai îndepărtate de centrul absolut al rețelei, cu atât ele sunt mai puțin populare și în mod automat mai excentrice, confirmând faptul că datele obținute în mod empiric reflectă statutul din lumea reală.

Astfel, comunitatea formată de genul muzical Pop/Rock este una strânsă și foarte centrală. Acest lucru se datorează faptului că muzicienii acestui gen muzical preferă să colaboreze doar între ei. Pe de altă parte, comunitățile Jazz și Blues au tendința de a se disipa și suprapune peste alte genuri muzicale. Acest lucru se datorează faptului că acești muzicieni sunt predispuși la colaborări cu artiști din alte genuri muzicale. Același lucru se poate spune și despre muzica Clasică, ce în cultura noastră actuală aduce contribuții semnificative la coloanele sonore ale filmelor, reclamelor, *etc.* În final, din genul Country, similar cu Pop/Rock, fac parte muzicieni care preferă să colaboreze între ei, mai bine decât cu muzicieni din alte genuri muzicale. Acest gen are însă o poziție mai excentrică în rețea, fapt datorat popularității scăzute.

În Figurile 4.3 - 4.6 putem observa distribuțiile de centralități în MuSeNet, specifice de altfel și altor rețele, și anume o distribuție de tip power-law, degree, eigenvector și pagerank. De remarcat este clusterul dominant din Figura 4.5, format din noduri cu un eigenvector mare. La o inspecție amanunțită, am determinat că această comunitate este formată din muzicieni maturi, cum ar fi Alphonso Johnson, cu studioul de înregistrări; faptul că întreaga muzică din rețea trece prin studiourile lor, îi determină să fie comunitatea centrală din rețeaua MuSeNet. În contextul meritocrație vs. topocrație, menționat anterior, această comunitate se folosește din plin de influența și meritele sale în această rețea topocratică, lucru valabil și din punct de vedere economic, dacă luăm în calcul următorul fapt: crearea și distribuirea muzicii este o activitate economică.

Întru-un final, similar cu studiul actorilor imdb, ce îl scoate în evidență pe actorul Kevin Bacon ca cel mai influent nod din rețea, am determinat următorul lucru: "Kevin Bacon"-ul industriei muzicale este Dave Grohl. Acest lucru este vizibil și în Figura 4.6, unde am aplicat metrica de pagerank, o metodă clasică de a determina noduri cu putere mare de influențare.

Graph Metric Analysis

Pentru a obține și alte rezultate concludente, am comparat rețeaua MuSeNet cu alte rețele distincte: Jazz și IMDB – datorită abordării lor similare –, dar și rețele de tip social networking, cum ar fi Facebook, Twitter sau Google+, pentru a pune în perspectivă proprietățile specifice artiștilor în comparație cu persoane de rând. Astfel, m-am folosit de metricile specifice rețelelor complexe, și anume: average degree, average path length, average clustering coefficient, modularity, graph edge density și graph diameter.

În mod interesant, rețeaua Facebook se situează la un nivel mediu din punct de vedere al sociabilității, pe când rețeaua actorilor IMDB este mai sociabilă, iar MuSeNet din contră, mai puțin sociabilă. Această diferență se poate explica în felul următor: utilizatorii platformelor de socializare (*i.e.* oameni de rând) interacționează și crează noi legături într-un ritm normal; actorii însă, depind de numărul filmelor/serialelor din care fac parte, la care casting-ul este mereu diferit. Acest lucru determină ca rețeaua lor să fie mai densă, dar și mai sociabilă. Pe de altă parte, muzicienii preferă să colaboreze în interiorul trupei din care fac parte, și nu neapărat cu alți muzicieni. Acest lucru face ca MuSeNet să fie o rețea mai puțin densă, bazată pe comunități și, într-un final, mai puțin sociabilă.

Motif Distribution

O abordare populară în Network Science pentru a analiza rețele complexe, este prin identificarea network motifurilor, adică sub-grafuri repetitive, formate dintr-un număr bine-definit de noduri, care sunt specifice doar rețelei respective. Pentru studiul rețelei MuSeNet, am luat în considerare doar cele cu dimensiunea de 3 noduri, datorită simplității detecției, a numărului mare de apariții în rețea, dar și a relevanței ce le oferă. Așadar, am determinat distribuția motifurilor pentru fiecare rețea empirică, folosind algoritmul FANMOD, unul din cei mai rapizi algoritmi de detecție. Ca rezultat am obținut Figura 4.7, în care putem observa că rețeaua muzicienilor Jazz, având o distribuție uniformă a motifurilor, se comportă ca o rețea de socializare, pe când rețelele IMDB și MuSeNet conțin motifuri specifice doar acestor rețele.

Pentru o veridicitate mai bună a acestor rezultate, am aplicat metrica de fidelitate pentru a compara distribuția de motifuri între ele. Rezultatele obținute sunt prezentate în Tabelul 4.5, unde rețelele sunt cu atât mai similare, cu cât valoarea tinde spre 1.

În cel de-al cincilea capitol prezint simulatorul socio-economic TrEcSim, și totodată rezultatele simulărilor obținute cu acesta. Trecând din domeniul muzical în cel economic, un aspect important de înțeles este motivul pentru care unii agenți sunt favorizați față de alții, sau ce tip de rețea ar fi mai potrivită pentru întreaga societate din care facem parte. Totuși, astfel de aspecte pot fi determinate fie analizând rezultate din trecut, fie prin estimări calitative, însă niciodată din observații în timp real, datorită faptului că rețelele economice sunt sisteme complexe non-lineare. O altă abordare pentru analiza sistemelor economice constă în simplificarea lor în modele matematice, ori simularea acestora cu simulatoare capabile de predicții comportamentale ale agenților.

Ca atare, am creat simulatorul economic TrEcSim, pentru a observa corelația dintre proprietățile topologice ale rețelelor și distribuția veniturilor din acestea. Așadar TrEcSim:

- Folosește un algoritm euristic personalizat, inspirat din modelul de interacțiune bazat pe toleranță.
- Are multiple setări pentru fiecare simulare în parte.
- Implementează teoriile recunoscute în mediul economic.
- Se folosește de caracteristici aplicabile în lumea reală: folosirea oricărei topologii pentru simulări, crearea și evoluția dinamică a agenților economici, a produselor, prețurilor, a investițiilor, *etc.*

În esență, TrEcSim permite analiza computațională a rețelelor economice colaborative și emergente la o scară macro, cu parametri setați la nivel micro.

Una din cele mai interesante studii ce a atras atenția comunității științifice, este modelul descris de autorii Borondo referitor la sistemele economice integrate, model pe care de altfel se bazează și TrEcSim. Analiza acestor sisteme este deosebit de importantă deoarece crearea de noi legături economice este costisitoare și de multe ori influențată de rețeaua socială ce co-există în ele. Așadar, într-un sistem economic real veniturile pot fi clasificate în funcție de sursa lor, și anume fie de la producător, fie de la middleman, omul intermediar dintr-o tranzacție. Astfel, autorii argumentează, că odată ce numărul legăturilor din rețea scade, rețeaua trece dintr-o stare meritocratică (echitabilă, unde venitul este stabilit de talentul individual al agentului) într-o stare topocratică (nedreaptă, unde poziția topologică a agentului economic determină venitul obținut). Modelul descris de autorii Borondo este însă unul simplificat, datorită abordărilor matematice, respectiv statistice folosite:

- Folosește o singură topologie (cea random).
- Roluri asigurate și neschimbătoare.
- Un singur tip de produs cu un preț fixat.
- Toată analiza se face pe parcursul unei singure iterații.

Toate aceste aspecte impactează realismul rețelei și al studiului. Așadar, la TrEcSim am îmbunătățit aceste aspecte: se folosește de orice sistem complex ca bază pentru simulări, fie creat ad-hoc, fie importat. Mai mult, evoluția agenților și a rolurilor este variabilă, iar un astfel de exemplu putem observa în Figura 5.11: producătorii produsului *Pr1* din partea stângă, pot fi și consumatori ai produsului *Pr2* din partea dreaptă, într-un alt ciclu de simulare. Similar, evoluția produselor, a calității și cantității acestora, cât și a prețului, sunt de asemenea variabile, iar venitul obținut de către agenți poate fi investit în diferite acțiuni.

Model Description

Modelul din spatele simulatorului, prezentat în Figura 5.10, se poate împărți în trei mari componente, și anume: inițializarea, faza tranzacțională și faza decizională. Pe când inițializarea se face o singură dată, celelalte două se execută pentru fiecare ciclu de simulare.

În faza de inițializare au loc patru procese menite de a crea instanța simulării, și anume: *createNetwork*, ce crează rețeaua agenților economici, bazată pe setările făcute sau datele importate; *createProducts*, ce definește produsele și atributele acestora; *createProductions*, ce definește producțiile; și, într-un final, procesul *createNeeds*, ce definește sub-setul de necesități pentru fiecare agent economic.

În faza tranzacțională, agenții economici identifică necesitățile actuale bazate pe factorul de importanță, calitatea, cantitatea și prețul produsului. Astfel, fiecare iterație nouă a fazei tranzacționale începe cu procesul *getBestProduction*, pentru a identifica cea mai bună opțiune de unde să procure produsele necesare, pe când procesul *getAffordableQuantity* determină cantitatea maximă ce poate fi obținută în funcție de calitatea și cantitatea produselor. Faza tranzacțională se încheie cu procesul *finalizeTransaction*, ce finalizează tranzacțiile în funcție de atributele menționate anterior, iar pentru determinarea venitului individual am folosit Ecuația 5.4, o variantă modificată a ecuației folosite și de către autorii Borondo.

Faza tranzacțională, deși importantă din punct de vedere economic, nu atrage după sine în mod automat dinamicitatea rețelei. Pe de altă parte, în faza decizională, agenții economici pot investi într-una din următoarele acțiuni:

- **Acțiunea 1:** crearea de noi legături între agenți.
- **Acțiunea 2:** crearea de noi produse.
- **Acțiunea 3:** îmbunătățirea produselor existente.
- **Acțiunea 4:** crearea de noi agenți economici.

Astfel, procesul *determineCurrentDecision* determină cea mai profitabilă investiție a agentului economic curent, după o analiză prealabilă a deciziilor luate în trecut, atât de agentul curent (*getPastDecisionScores*), cât și de la un număr arbitrar de vecini (*getPastDecisionScoresFromNeighbours*). Odată determinată acțiunea, procesul *makeDecision*, ultimul proces din

această fază, implementează acțiunea.

Simulation Results

Pentru a obține rezultate concludente, am folosit aceleași setări inițiale la toate simulările mele. Ca atare, am generat 10 rețele a 100,000 agenți fiecare și cu densități diferite pentru fiecare topologie de bază: mesh, small-world, random și scale-free. Astfel, rețelele obținute sunt prezentate în Tabelul 5.3, și sunt create pentru a simula două cazuri pentru fiecare topologie în parte: unul în care rolul de producător este asignat în mod aleatoriu agenților economici, iar unul în care algoritmul asignează rolurile cu o anumită probabilitate, în funcție de degree-ul agentului: cu cât e mai interconectat, cu atât probabilitatea pentru a fi producător va fi mai mare.

Simulation Results for the Extended Model

Simulând venitul cumulativ al agenților economici pentru primul caz, atunci când producătorii sunt asignați în mod aleatoriu, am obținut graficele din Figurile 5.14 - 5.16, din care reiese dificultatea producătorilor de a depăși venitul intermediarilor. Astfel producătorii depășesc pragul de 50% abia după ~230 de iterații. Pe de altă parte, în cazul rețelei scale-free (Figura 5.17), producătorii nu reușesc să atingă nici măcar acest prag. Acest lucru se datorează prezenței hub-urilor guvernate de către intermediari. Așadar, mărinind doar densitatea rețelelor, nu se garantează și prezența meritocrației.

Pe de altă parte, asignând rolurile de producători în mod preferențial agenților cu degree-uri mari, obțin o tranziție evidentă, de la o rețea topocratică la una meritocratică. Astfel, pentru toate cele 4 topologii, producătorii depășesc veniturile cumulate ale intermediarilor în ~150 iterații, ~200 în cazul topologiei scale-free. Totuși, un astfel de scenariu nu este unul realist, datorită faptului că rețelele din lumea reală nu vin cu astfel de condiții.

Simulation Results for the Distribution Payoff

Pentru a obține distribuția veniturilor pentru fiecare rol economic, m-am folosit de datele obținute anterior și am analizat atent distribuțiile între cele două categorii: producători și intermediari, obținând Figurile 5.22 - 5.25. Atunci când producătorii sunt asignați în mod aleatoriu, venitul mediu al producătorilor este reprezentat de o distribuție pozitiv înclinată (*i.e.* positively skewed distribution); cu alte cuvinte, doar câțiva agenți beneficiază de un venit mărit față de restul agenților producători. Pe de altă parte, distribuția agenților intermediari se aseamănă cu o distribuție Gaussiană, dovada clară că există mai mulți intermediari care obțin un venit mărit, atât comparativ cu producătorii, cât și cu restul intermediarilor. În cazul rețelei scale-free, venitul producătorilor se aseamănă unei distribuții de tip log-normal; acest lucru se datorează prezenței huburilor, ce avantajează majoritatea intermediarilor.

Asignând rolurile de producători în mod preferențial agenților cu un degree mare, obțin o distribuție Gaussiană pentru veniturile producătorilor, respectiv o distribuție pozitiv înclinată pentru intermediari, lucru vizibil și în Figurile 5.26 - 5.28. Faptul că aceste două distribuții alternează în funcție de asignarea rolurilor de producători, ne demonstrează atât importanța, cât și consecința acestui aspect.

Și de această dată rețeaua scale-free (Figura 5.29) oferă rezultate interesante: producătorii aflați în locații-cheie câștigă foarte mult, semnificativ mai mult decât intermediarii sau restul producătorilor, obținând o distribuție de tip power-law.

Simulation Results Pertaining the Ergodicity in Economic Networks

După identificarea distribuțiilor, am obținut datele din Tabele 5.5 - 5.8. Acest proces de identificare (*i.e.* fitting) a fost făcut folosind aplicația EasyFit, o aplicație folosită pentru fitting în sisteme dinamice. Analiza acestor date ne arată că există o oarecare similitudine între distribuțiile în timp și numărul agenților economici. Am obținut diferențe absolute între 2% și 27% pentru rețelele mesh (Figurile 5.30 și 5.31), small-world (Figurile 5.32 și 5.33) și random (Figurile 5.34 și 5.35), atât la asignarea aleatorie cât și preferențială a producătorilor, demonstrându-

ne că aceste topologii sunt într-adevăr ergodice. La modul general, doar rețeaua scale-free (Figurile 5.36 și 5.37) conține diferențe sesizabile, lucru evidențiat atât din analiza grafică, dar în același timp și din cea analitică, diferența absolută fiind între 22% și 80%.

În ultimul capitol menționez concluziile relevante din această teză. Astfel, rețelele complexe sunt studiate exhaustiv datorită aplicației lor în diferite domenii, printre care și ingineria sau computer science. Folosind puterea analitică a calculatorului, putem genera, simula și analiza rețele complexe într-o manieră dinamică, pentru a identifica anumite modele și relații existente. În această lucrare am analizat două astfel de rețele, și anume MuSeNet, o rețea naturală, cât și rețele economice sintetice create cu simulatorul TrEcSim.

Analizând MuSeNet în prima parte a tezei, am obținut informații pertinente despre rețeaua colaborativă a muzicienilor; este o rețea scale-free, cu o distribuție a muzicienilor sub formă de power-law și cu un grad mare de centralitate. Este mai puțin sociabil decât rețeaua actorilor, datorită felului în care muzicienii colaborează între ei, și din același motiv legăturile se formează într-un ritm mult mai lent față de rețele de socializare. Am aplicat o metodă nouă de analiză a rețelelor prin expresii numerice, bazată pe distribuția motifurilor, și am identificat artiști cu un grad mare de centralitate – *e.g.* Dave Grohl, “Kevin Bacon”-ul industriei muzicale –, dar și artiști cu un degree și pagerank mare. De asemenea, am evidențiat prezența unei comunități dominante, formată din artiști cum ar fi Alphonso Johnson, cu un eigenvector mare.

În partea a doua a tezei, am abordat simulatorul socio-economic TrEcSim și rezultatele obținute cu acesta. Astfel, creat pentru a putea simula rețele economice într-o manieră cât mai realistă, TrEcSim a fost folosit pentru a analiza distribuțiile de venit la scară macro, având reguli definite la scară micro. Folosind TrEcSim, am arătat că plasarea fizică a agenților economici influențează în mod direct distribuțiile de venit atât pentru producători, cât și pentru intermediari. De asemenea, am arătat și faptul că distribuțiile de venit din categorii similare urmează o curbă de tip power-law, și sunt influențate doar de plasarea agenților în rețea, nu și de topologia rețelei în sine. Mai mult, folosind o metodă nouă de analiză a evoluției veniturilor, am confirmat avantajul nedrept al rețelelor economice topocratice, față de cele meritocratice, indiferent de topologia de bază a rețelei respective, iar analizând ergodicitatea sistemelor în timp, cât și la nivel de populație, am concluzionat că distribuția de venit este într-adevăr ergodică pentru toate topologiile de rețele, mai puțin cel scale-free, fără ca rolul agentului economic să influențeze într-un fel sau altul acest aspect; un bun indicator de ergodicitate este și prezența unui număr (limitat) de agenți economici falimentați, aspect inexistent în modelul descris de autorii Borondo.

Indiscutabil, contribuțiile aduse cu această teză sunt semnificative în contextul folosirii uneltelor computaționale și algoritmilor propuși în domeniul Network Science – un domeniu de mare interes –, atât ca rezultate, cât și ca metodologii aplicabile la alte studii similare. De asemenea, toate analizele și metodologiile aplicate, se pot desfășura în continuare pentru obținerea de date noi.

Direcții de cercetare

Obținerea de rezultate relevante este forța motrice pentru orice cercetător, cu atât mai mult când domeniul în care activează este încă relativ tânără. Ca atare, având în vedere progresele recente în domeniul Social Network Analysis și direcția pe care studiile mele doctorale m-au dus, prevăd următoarele contribuții pentru a avea efect imediat asupra cercetării pe care am început-o:

- Algoritm euristic îmbunătățit: se va depune efort suplimentar pentru îmbunătățirea algoritmului euristic al lui TrEcSim. În prezent, algoritmul analizează deciziile anterioare luate de către agenții economici și calculează rezultatul acesteia; cu toate acestea, va fi îmbunătățit pentru a permite algoritmului euristic să utilizeze aceste

informații și să creeze o simulare tampon. Cu alte cuvinte, algoritmul va crea o simulare secundară în avans și își va analiza rezultatele în mod probabilistic, îmbunătățind considerabil acuratețea alegerilor adoptate.

- Algoritm genetic: conversia de la algoritmul euristic existent la unul genetic ar îmbunătăți TrEcSim în mai multe moduri. Implementarea acestui tip de algoritm ar permite utilizatorilor să găsească soluții potrivite într-un timp de calcul scurt, în timp ce o mutație aleatorie ar garanta o gamă mai largă de soluții.
- Teorii economice: o îmbunătățire a implementării principalelor abordări economice va spori considerabil aplicabilitatea și utilizabilitatea simulatorului TrEcSim în domeniul analizei rețelelor complexe (emergente), inclusiv din domeniul economic; două astfel de teorii sunt: „the theory of marginality” și „the labor theory of values”.
- Realism: prin implementarea a noi mecanisme în TrEcSim, se va îmbunătăți, fără îndoială, realismul simulatorului, luând în considerare mai mulți factori din lumea reală, ca de exemplu asimetria informațională - care apare adesea în tranzacții – sau implicarea guvernului și a reglementării. Adăugarea costurilor (sau a unei alte forme de consum pentru agentul economic) în menținerea anumitor acțiuni în vigoare (*e.g.* legături, produse noi, producție îmbunătățită *etc.*) va contribui de asemenea la realismul menționat.
- Interfață: este necesară o interfață și mai personalizabilă pentru îmbunătățirea interacțiunii cu utilizatorul.
- Simulări ample: simulările din această teză reprezintă doar câteva scenarii posibile pe care le putem analiza într-o manieră realistă folosind simulatorul TrEcSim. În consecință, continuarea cercetării și simularea sistemelor economice din lumea reală prin utilizarea altor setări posibile de configurare în cadrul TrEcSim poate produce rezultate semnificative, cum ar fi, de exemplu, saturația și deficitul de produs.

Bibliografie selectivă:

1. Acemoğlu, D., Como, G., Fagnani, F., and Ozdaglar, A. (2013). Opinion fluctuations and disagreement in social networks. *Mathematics of Operations Research*, 38(1):1–27
2. Alberich, R., Miro-Julia, J., and Rosselló, F. (2002). Marvel universe looks almost like areal social network. *arXiv preprint cond-mat/0202174*
3. Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47
4. Barina, G., Sicoe, C., Udrescu, M., and Vladutiu, M. (2017). Simulating trade in economic networks with trecsim. In *European Network Intelligence Conference*, pages 169–185. Springer
5. Barina, G., Topirceanu, A., and Udrescu, M. (2014). Musenet: Natural patterns in the music artists industry. In *2014 IEEE 9th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI)*, pages 317–322. IEEE.
6. Barina, G., Udrescu, M., Barina, A., Topirceanu, A., and Vladutiu, M. (2019). Agent based simulations of payoff distribution in economic networks. *Social Network Analysis and Mining*, 9(1):63
7. Barina, G., Udrescu, M., Topirceanu, A., and Vladutiu, M. (2018). Simulating payoff distribution in networks of economic agents. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 467–470. IEEE
8. Blume, L. E. and Durlauf, S. N. (2005). *The economy as an evolving complex system, III: Current perspectives and future directions*. Oxford University Press
9. Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). Network analysis in the social sciences. *science*, 323(5916):892–895

10. Borondo, J., Borondo, F., Rodriguez-Sickert, C., and Hidalgo, C. A. (2014). To each according to its degree: The meritocracy and topocracy of embedded markets. *Scientific reports*, 4:3784
11. Chen, D., Lü, L., Shang, M.-S., Zhang, Y.-C., and Zhou, T. (2012). Identifying influential nodes in complex networks. *Physica a: Statistical mechanics and its applications*, 391(4):1777–1787
12. Dawid, H., Gemkow, S., Harting, P., van der Hoog, S., and Neugart, M. (2014). Agent-based macroeconomic modeling and policy analysis: the eurace@ unibi model
13. Duma, A. and Topirceanu, A. (2014). A network motif based approach for classifying online social networks. In *Applied computational intelligence and informatics (SACI), 2014 IEEE 9th international symposium on*, pages 311–315. IEEE
14. e Fonseca, B. d. P. F., Sampaio, R. B., de Araújo Fonseca, M. V., and Zicker, F. (2016). Co-authorship network analysis in health research: method and potential use. *HealthResearch Policy and Systems*, 14(1):34
15. Fass, C., Ginelli, M., and Turtle, B. (1996). Six degrees of kevin bacon. Plume Books
16. Franceschet, M. (2011). Collaboration in computer science: A network science approach. *Journal of the American Society for Information Science and Technology*, 62(10):1992–2012
17. Freeman, L. (2004). The development of social network analysis. *A Study in the Sociology of Science*, 1
18. Giaquinto, G., Bledsoe, C., and McGuirk, B. (2007). Influence and similarity between contemporary jazz artists, plus six degrees of kind of blue. Ph. D. Dissertation, Citeseer.
19. Kantarci, B. and Labatut, V. (2013). Classification of complex networks based on topological properties. In *Cloud and Green Computing (CGC), 2013 Third International Conference on*, pages 297–304. IEEE
20. Matthews, L. and Richard, P. (2017). Who is central to a social network? it depends on your centrality measure
21. Peters, O. (2019). The ergodicity problem in economics. *Nature Physics*, 15(12):1216–1221
22. Staudt, C., Schumm, A., Meyerhenke, H., Görke, R., and Wagner, D. (2012). Static and dynamic aspects of scientific collaboration networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, pages 522–526. IEEE
23. Topirceanu, A., Barina, G., and Udrescu, M. (2014a). Musenet: Collaboration in the music artists industry. In *2014 European Network Intelligence Conference*, pages 89–94. IEEE
24. Topirceanu, A., Udrescu, M., and Vladutiu, M. (2013). Network fidelity: A metric to quantify the similarity and realism of complex networks. In *Cloud and Green Computing (CGC), 2013 Third International Conference on*, pages 289–296. IEEE
25. Topirceanu, A., Udrescu, M., Vladutiu, M., and Marculescu, R. (2016). Tolerance-based interaction: a new model targeting opinion formation and diffusion in social networks. *PeerJ Computer Science*, 2:e42
26. Wernicke, S. (2006). Efficient detection of network motifs. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 3(4)