

Rezumat al tezei de doctorat

Analiza și modelarea structurală și comportamentală a societății

Structural and Behavioral Analysis and Modeling of the Society

ing. Alexandru Topîrceanu

Susținută la data de 12.Februarie.2016

Sunt doctorand în cadrul Departamentului de Calculatoare și Tehnologia Informației, din Facultatea de Automatică și Calculatoare a Universității Politehnica Timișoara, aflat sub conducerea d-lui prof. dr. ing. Mircea Vlăduțiu. Teza mea de doctorat se intitulează “Structural and behavioral analysis and modeling of the society” (*Analiza și modelarea structurală și comportamentală a societății*) și a fost elaborată în perioada octombrie 2012 - noiembrie 2015, fiind în momentul de față finalizată și aprobată de comisa de conducere de doctorat pentru susținerea publică și acordarea titlului de doctor în știința calculatoarelor.

Această teză este elaborată din perspectiva științei calculatoarelor pentru a aduce contribuții originale și substanțiale domeniului de analiză de rețele sociale (social networks analysis), o arie de mare interes științific recent, sub egida mai marelui domeniu de rețele complexe, numit și *New Network Science*.

Nouă știință a rețelelor (*New Network Science*) capătă un interes crescut din mai multe domenii ale științei, deoarece multe observații empirice ale lumii noastre înconjurătoare, a fenomenelor naturale, a proceselor realizate de om, arată aceleași proprietăți funcționale, indiferent dacă acestea sunt de origine naturală sau sintetică. Există, spre exemplu modele de rețele complexe care descriu interacțiuni tectonice și planetare, interacțiuni gastronomice între rețete culinare, rețelele neuronale ale creierului, mecanismele de interacțiune dintre bacterii, utilizarea mijloacelor de transport în orașe, rețele de citare și colaborare științifică, internetul, rețelele de distribuție a energiei electrice, relații ale cuvintelor într-o limbă, procesul decizional social, al structurilor politice etc. Toate acestea pot fi modelate și înțelese prin rețele complexe.

O importantă ramură a acestei științe este *Analiza de rețele sociale* care își propune să înțeleagă și să modeleze rețele sociale reale. Această știință se poate concentra pe nivelul topologic de rețea, adică asupra modului de interconectare a nodurilor, sau pe nivelul comportamental, adică modul în care nodurile interacționează. Ambele aceste probleme sunt abordate prin intermediul unor studii formale, validate prin date empirice (măsurători directe și indirecte, sondaje, analize

statistice etc.), care duc spre o îmbunătățire treptată a modelelor de rețele sociale din lumea reală observată. Interdisciplinaritatea științei rețelelor sociale reunește mai multe domenii ale științei care prelucrează cantități uriașe de date (*big data*) modelate sub formă de grafuri matematice.

Structură și contribuții științifice

Teza este redactată în limba engleză, este structurată pe 7 capitole, de la introducere la concluzii, se întinde pe 229 pagini, și are de asemenea 4 anexe importante și originale care susțin calculul și interpretările rezultatelor din teză. Teza conține 38 de tabele și 80 de figuri complexe. Lucrarea se încheie cu o lista bibliografică impresionantă de 297 referințe relevante domeniului, dar și lucrări fundamentale în știință. Toate aceste date susțin în continuare buna pregătire teoretică a doctorandului și calitatea tezei sale.

Activitatea din perioada cercetării doctorale este reprezentată prin 6 articole de jurnal cu factor de impact, apoi 5 lucrări de conferințe indexate ISI și 5 lucrări de conferințe indexate IEEE (10 în total). Dintre acestea, se așteaptă ca 4 din lucrările IEEE să se indexeze și ISI în viitorul apropiat; de asemenea doctorandul este în așteptarea indexării a încă 5 lucrări acceptate și prezentate la conferințe internaționale. Doctorandul face parte și din două proiecte de cercetare câștigate, unul finanțat de firma *Linde*, și unul național de tip PN-II *Tinere Echipe*.

Domeniul tezei de doctorat

Această teză abordează domeniul de analiză de rețele sociale, oferind contribuții originale din perspectiva științei calculatoarelor. Pornind de la dezideratele sus-menționate, o direcție notabilă a cercetării științifice actuale este studiul și înțelegerea dinamicii opiniei sociale și fluctuațiile de opinie individuală, a modului în care oamenii se influențează reciproc și modul în care toate aceste elemente pot fi prezise și controlate. Beneficiul de a înțelege procesele complexe din spatele modul în care oamenii adoptă și își formează propriile opinii despre problemele înconjurătoare este o preocupare majoră pentru științe cum ar fi psihologia, filosofia, politica, marketingul, științele financiare etc.

Politica, spre exemplu, se folosește de studii pentru a estima influența partidelor politice și mijloacele pentru a crea un consens în rândul alegătorilor. Fie că se urmărește un acord, un consens de cooperare sau colaborare, partidele politice sunt interesate de opinia publică generală, mai degrabă decât de opinia indivizilor. Științele financiare, pe de altă parte, studiază piețele și consumatorii pentru a îmbunătăți profiturile. Marketingul folosește mai multe tehnici pentru a înțelege nevoile, punctele forte și punctele slabe ale diferitelor straturi sociale sau grupuri. Astfel, o similitudine cu rețelele de calculatoare poate fi observată, în care *throughput*-ul (debitul) și

livrarea corectă de pachete sunt importante, mai degrabă decât performanța și specificațiile unui nod individual.

Domeniile științifice tratate în această teză sunt: analiza rețelelor sociale, rețele complexe, topologii de rețea, dinamica de opinie socială, teoria grafurilor și statistica.

Conform studiului literaturii, nevoile de interes major în *analiza rețelelor sociale* în 2015/2016 sunt: (1) nevoia unor modele topologice de interconectare socială mai realiste, (2) nevoia unor modele de dinamică topologică mai fiabile, (3) nevoia de rafinare a modelelor de interacțiune socială, (4) nevoia de modele mai fiabile în ceea ce privește acuratețea predicției.

În conformitate cu aceste direcții revendicate de lumea științifică, teza de doctorat se axează pe două direcții principale, la fel de importante pentru modelarea opiniei sociale. Rezumând primele două nevoi menționate, teza propune crearea de rețele ce modelează și urmăresc dinamic tiparele de conectare ale indivizilor unei societăți reale (*cercetarea structurală*). Rezumând ultimele două nevoi menționate, teza prezintă un model robust de interacțiune socială, și anume un set de reguli care descriu modul în care agenții se comporta într-o societate (*cercetarea comportamentală*).

Ambele direcții – cea structurală și cea comportamentală – sunt importante deoarece topologia creează mijloacele de interconectare, în timp ce modelul de comunicare definește evoluția reală a procesului decizional al indivizilor. Pentru aceasta, au fost cercetate minuțios ambele direcții în literatura de specialitate pentru această teză.

Rezumatul tezei pe capitole

Lucrarea începe cu un capitol de **introducere** în care este prezentat domeniul tezei de doctorat, cel al rețelelor complexe și al analizei rețelelor sociale. Aici se subliniază direcțiile existente în aria științifică, și se aduc de asemenea referințe din domenii adiacente, precum psihologie și psihologie socială. Analiza de rețele sociale ajută la înțelegerea mecanismelor care transformă opinia individului într-o opinie răspândită la nivel social; a felului în care opiniile afectează indivizii și modul în care aceștia evoluează în timp așa cum se vede la scară microscopică. Modelarea comportamentului social poate fi atât un mijloc de apărare cât și de creștere al drepturilor democratice, precum și un mijloc de impunere sau manipulare într-o societate sau pe un strat social, iar toate aceste evenimente sunt rezultatul interconectării sociale și evoluției opiniei sociale între oameni.

În contextul tezei prezentate, contribuțiile se axează pe modele predictive bazate pe procesele de modelare pentru difuzie, și utilizează rețele empirice (Facebook, Twitter, Google Plus, MemeTracker, Yelp) pentru scopuri de validare. De asemenea, se oferă o perspectivă topologică a modului în care structura de rețea socială apare și evoluează în timp. Acest lucru este realizat prin crearea unui model topologic original, prin introducerea unui model de evoluție structurală a

rețelelor, prin definirea și validarea unui model de interacțiune socială, și prin introducerea unei metodologii noi de analiză a rețelelor sociale empirice folosind motive de rețea (*network motifs*).

Motivația tezei este susținută în printr-o listă de referințe de specialitate care redau direcțiile de cercetare necesare în viitorul apropiat. Prima provocare este de a determina adevărata "opinie" în societate, acele cuvinte cheie care sunt circulate între indivizi. Media socială este un mijloc de exprimare general valabil al vieții noastre de zi cu zi, și conține o mare parte a comunicării umane în secolul 21, astfel că devine foarte important să fim capabili să filtrăm aceste subiecte de interes pentru un anumit context de cercetare (de exemplu, doar cercetare de piață, sau context politic etc.). După acest pas, cercetarea poate fi orientată spre înțelegerea și modelarea proceselor de difuzie ale acestor opinii. În cele din urmă, odată ce opinia și difuzia sunt cunoscute, este important să se determine care noduri acționează ca distribuitoare ale opiniei. Acest lucru ajută la simularea și precizarea rezultatului fenomenelor de difuzie.

În conformitate cu domeniul de doctorat, capitolul al doilea de **fundamentare teoretică** începe cu o secțiune numită „*Rețele sociale: o introducere în știința calculatoarelor*” care este menită să fundamenteze rolul esențial pe care îl are știința calculatoarelor în această arie inter-disciplinară. Se oferă un punct de început al înțelegerii tezei atât pentru un inginer în calculatoare la modul general, cât și pentru un specialist în știința rețelelor. Numărul mare de referințe științifice care acompaniază majoritatea paragrafelor, dau rigoare acestui capitol care este extrem de bine fundamentat și convingător pentru orice cititor.

Cu excepția unor concepte comune și fundamentale ale teoriei grafurilor, acest capitol, și teza în general, se concentrează exclusiv pe noțiuni specifice contextului social. După ce este introdus cititorul relației dintre informatică și analiza de rețele sociale, capitolul prezintă metricile de grafuri cele mai relevante, utilizate în știință. Se explică noțiunile de noduri, muchii, ponderi, distribuții de centralitate, modularitate, structură de comunitate, drumuri minime etc., urmate de concepte specific sociale care sunt folosite pentru modelarea opiniei. Cea de a patra subsecțiune este dedicată topologiilor de rețea socială, care reprezintă una dintre cele două puncte principale de studiu pentru această teză. Aici se prezintă modelele fundamentale care au inspirat crearea unor topologii mai avansate, care au fost o sursă de inspirație pentru propunerile originale elaborate în această teză. Apoi, cu referire la partea a doua a tezei, se introduc detaliat modele state-of-the-art de interacțiune socială. Capitolul se încheie printr-o discuție despre standardele, tendințele și limitările în crearea și compararea modelelor sociale.

Capitolul al treilea este intitulat **Modelarea datelor empirice folosind grafuri**, și reprezintă un set introductiv de contribuții originale care stau la baza modelării topologice și comportamentale din această teză. În alte cuvinte, prin diversitatea și complexitatea lor, aceste studii au facilitat un punct de plecare fundamental și o experiență valoroasă doctorandului. Pentru a putea defini parametrii pentru topologii realiste, și să se înțeleagă principiile de interacțiune socială, s-au făcut mai multe studii bazate pe seturi de date empirice, toate prezentate și publicate

în lucrări de conferințe și reviste de calculatoare și medicale. Pentru a discerne care sunt proprietățile reprezentative ale rețelelor sociale, s-au făcut două studii privind datele sociale empirice care reprezintă rezultate noi, în premieră în literatura prin perspectiva oferită. Primul model de studiu reprezintă rețeaua de tuturor artiștilor din industria muzicală, intitulată MuSeNet (susținută prin două lucrări de conferințe, cu indexare ISI). În al doilea studiu, s-a construit o rețea pentru lumea modei numită FMNet (o lucrare la o conferință IEEE prestigioasă). Abordarea inovatoare din spatele acestui al doilea studiu i-a adus doctorandului și premiul de „*Best Paper Award*” la conferința cu participanți de prestigiu din Europa la care a luat parte doctorandul. Ambele topologii au fost generate și analizate cu ajutorul tehnicilor state-of-the-art din literatura analizei de rețele complexe, și folosind unealta Gephi împreună cu cod Java dezvoltat personal pe platforma Netbeans.

Împreună cu aceste proiecte, doctorandul a colaborat și în colectivul de cercetare ACSA, unde și-a propus să analizeze structura formată de rețelele rutiere din orașe dintr-o perspectivă inovatoare. În colectiv, și inspirat de abordări similare de comparare a rețelelor, s-a creat o metodologie care propune evaluarea rețelelor rutiere ale unui oraș pe baza distribuției de motive de rețea. Aceste colaborări s-au materializat printr-un număr de 5 de lucrări de proceedings IEEE.

Totodată, doctorandul s-a implicat în cercetare medicală, și anume modelarea bazată pe *network medicine*, unde a obținut o serie de publicații, dintre care cele mai notabile rezultate sunt obținute în modelarea rețelei de compatibilitate de risc între pacienții cu apnee de somn (~10 lucrări la congrese medicale, din care multe indexate în reviste medicale. Două lucrări sunt indexate ISI și sunt publicate în jurnalele *European Respiratory Journal*, respectiv *Chest*, ambele cu factor de impact peste 7). De asemenea, s-a mai întreprins și cercetare în modelarea rețelei de compatibilitate a pacienților cu boli de inimă (o lucrare de jurnal, în *Irish Journal of Medicine*). Aceste studii inter-disciplinare au adus îmbunătățiri substanțiale de predicție și diagnoză a pacienților. Nu în ultimul rând, s-a propus o abordare total originală folosind motive de rețea (*network motifs*) pentru a extrage caracteristici topologice a rețelelor sociale online (lucrare de jurnal în Elsevier *Computer Communications*).

Capitolul patru se referă la **Generarea realistă a topologiilor de rețele sociale**, și prezintă contribuția originală amănunțită și validată pe direcția de modelare structurală revendicată în teza de doctorat. Se începe prin a motiva faptul că principalele direcții în analiza de rețele sociale (social networks analysis) sunt înțelegerea și definirea de modele topologice mai realiste. Un obiectiv al lumii științifice este de a crea un model al societății umane, cu toate acestea, complexitatea unui astfel de model este crescută de natura interacțiunii umane, iar studiile actuale nu reușesc să creeze un model complet realist a societăților în care trăim. Abordarea propusă este inspirată din studiile de rețele sociale online, rețele de colaborare din reale din industria muzicală și a modei, studii în medicină care implică modelarea pacienților ca grafuri, și alte studii. Se combină aceste perspective cu puterea algoritmilor genetici de a optimiza date topologice într-un mod natural. Acest lucru duce la modelul original numit Genosian, care combină proprietățile small-world și scale-free

pentru a crea o rețea socială bazată pe comunități, care este apoi rearanjată folosind datele obținute empiric de la rețelele de prietenie de Facebook, și optimizată folosind algoritmi genetici.

Lucrarea, care este publicată ca un capitol de carte Springer indexat ISI, urmărește potrivirea unui număr de șase metrici de graf (gradul mediu al nodurilor, drumul mediu în graf, coeficientul de clusterizare, modularitatea, diametrul și densitatea grafului) într-un mod optim euristic, astfel încât acestea să se regăsească în niște intervale specifice, reprezentative, descoperite prin studii pe date reale. Aliniind toate aceste șase metrici este un lucru improbabil de realizat determinist, datorită complexității structurale obținute. De aceea, doctorandul, dând dovada de mare pricepere inginerescă, se folosește de algoritmi genetici atribuind un *fitness* specific fiecărei muchii ce leagă noduri din două comunități diferite. Urmărind optimizarea fitness-ului tuturor muchiilor din graf, aceste muchii se desfac și se reconectează la alte noduri. Ca urmare, topologiile de rețele sociale sintetice obținute sunt mai realiste, cu o fidelitate cu 63% mai aproape de parametrii observați în lumea reală decât cel mai apropiat competitor din cinci alte topologii analizate. Această evaluare se bazează pe setul definit de șase metrici de grafuri care sunt considerate în literatura de specialitate ca fiind reprezentative pentru caracterizarea structurii de rețea.

Capitolul cinci, intitulat **Betweenness ca forță din spatele emergenței și evoluției rețelelor sociale**, aduce o propunere matematică originală, cu o nuanță filozofică, care explică și motivează nevoia de evoluție a rețelelor sociale în care trăim, abordând chiar la concepte din psihologie socială. Cercetarea actuală în domeniul rețelelor sociale se concentrează pe construirea de modele topologice care reflectă structura și evoluția acestor rețele complexe. Cum fenomenul de emergență al rețelelor sociale este, în general, condus de reguli simple de atașament preferențial, cercetarea din acest capitol își propune să descopere care centralitate de graf este cea mai importantă în definirea preferințelor, care duc la rândul lor la conexiuni noi în graf. Bazat pe literatura de specialitate recentă și propria analiză empirică a doctorandului, se dovedește, prin măsurători, că centralitatea betweenness a nodurilor este principala forță care conduce apariția rețelelor sociale. Acest articol, în prelucrare la revista *PLoS One*, încearcă să răspundă la o întrebare fundamentală: care este ingredientul principal pentru crearea de legături noi într-o rețea socială? Pentru a aborda în mod consecvent această întrebare s-a analizat topologia, deoarece reflectă modul în care rețeaua a fost structurată în cursul evoluției sale.

Se obțin rezultate sintetice prin crearea de modele topologice folosind mai multe centralități relevante: degree, eigenvector, closeness, betweenness și clustering, și combinații ale acestora luate câte două și câte trei. În toate cazurile, rețelele ce conțin betweenness sunt mai fidele față de modelele empirice. Se studiază și modelează de asemenea și grafuri ponderate, un deziderat foarte râvnit de literatura de specialitate, și reușit în această teză. Urmând o serie de indicii complete din literatura de specialitate, precum și prin analiza seturilor de date din lumea reală, se concluzionează cum că centralitatea betweenness este principala forță motrice din spatele formării

de rețea sociale și a evoluției acestora. Toate demersurile prezentate în acest capitol sunt validate, iar metodologia este robustă.

Capitolul șase, intitulat **Interacțiune bazată pe toleranță: un nou model pentru formarea opiniei și difuziei în rețele sociale**, reprezintă contribuția pe cea de a doua direcție a tezei, și anume modelarea comportamentală a societății. Lucrarea a fost acceptată la revista *PeerJ Computer Science*, o revistă nouă dar pretențioasă, inspirată după modelul *PLoS One*, și care oferă un proces riguros de peer-review oferit de cercetători renumiți din întreaga lume. Una din principalele motivații din spatele analizei de rețele sociale este înțelegerea formării opiniei și a difuziei de informație. Modelele anterioare au limitări, deoarece în mod specific presupun mecanisme de interacțiune de opinie pe baza unor praguri, fie fixe, fie care evoluează conform unui proces aleatoriu. Într-adevăr, analiza empirică, propusă în teză, pe seturi de date reale de mari dimensiuni, cum ar fi Twitter, MemeTracker, și Yelp, descoperă un set de fenomene dinamice la nivel de populație nedescoperite anterior, și anume existența unor faze distincte de formare de opinie și de echilibrare socială. De asemenea, se arată că o tranziție de fază de la un comportament neregulat (și nepredictibil) la echilibrare socială poate fi declanșat de topologie și de concentrația de surse de opinie. Prin urmare, pentru a construi un model care reprezintă în mod corespunzător aceste fenomene, se propune un nou model de interacțiune (la nivel individual) bazat pe *toleranță*. Spre deosebire de modelele de interacțiune de opinie existente, modelul de toleranță presupune dorința interioară a individului de a accepta noi opinii, iar aceasta evoluează în timp în funcție de trăsăturile de bază ale omului.

Introducerea conceptului de toleranță se face prin ecuații matematice ce susțin observațiile empirice, și sunt fundamentate pe considerente din psihologia socială. Pornind de la observațiile pe seturile de date din Twitter, Yelp și MemeTracker, se trece la simulare cu evenimente discrete pe diverse topologii de rețea socială. Modelul original de interacțiune socială este validat matematic, prin simulare, și cu date reale, și arată că mărimea rețelei și concentrația de surse de opinie sunt importante, iar tranziția de fază la echilibrarea socială este favorizată în principal de structura democratică de topologii *small-world*.

Studiul dinamicii de opinie prin conceptul propus de echilibrare socială arată caracteristici cheie care pot fi utilizate în aplicații practice, cum ar fi înțelegerea nevoilor unor utilizatori/clienti, marketing, rezolvarea conflictelor. Sub cerința de a menține echilibrul social, și de a nu se ajunge la starea de intoleranță, lucrarea oferă o clasificare a topologiilor de rețea bazată pe proprietatea de echilibrare socială. Rețelele cu structura *small-world* sunt de tip democratic și promovează comunicare și echilibru; fenomenul este, de asemenea, expus în cazul în care există o concentrație mare de agenți (surse) de opinie care stabilizează opinia în rețelele de tip *mesh*. Dacă există semnificativ mai puține surse de opinie în rețea, echilibrarea se va realiza numai în cazul în care se folosește o strategie de plasare pentru a contracara opiniile rivale. Pe de altă parte, topologia *scale-free* este de tip oligarhic, și prezintă un avantaj clar al agenților plasați strategic în noduri centrale sau pe trepte superioare ierarhice, care fac nodurile din nivelurile inferioare ale arborelui lipsite de

putere. Fenomenul de echilibrare nu se reproduce în rețelele cu proprietăți *scale-free*. În mod evident, conceptul de echilibrare socială reprezintă un concept valoros și original, și deschide o nouă direcție de dezbătut în literatura în anii ce vor urma.

În al șaptelea capitol, cel de **concluzii**, se prezintă o serie valoroasă de contribuții personale pentru domeniul analizei de rețele sociale. Eu fac parte din două proiecte de cercetare, unul pe tematică medicală, în echipa Morpheus, unde s-au obținut deja o pletoară impresionantă de publicații (din care 2 articole de jurnal cu factor de impact peste 7), și un al doilea în echipa NOVAMOOOC, pe tematică educațională și de analiză de rețele sociale, unde îmi voi folosi capacitatea analitică dobândită în cadrul anilor de cercetare ca doctorand. Per total, dețin 6 articole de jurnal cu factor de impact, un capitol de carte ISI, patru lucrări ISI la momentul de față (încă 7 lucrări ce urmează a fi indexate ISI), și 5 lucrări IEEE. De menționat este și colaborarea cu grupul de cercetare de la Carnegie Mellon (SUA), condus de prof. Mărculescu, fructificată cu o lucrare de revistă și multe noi idei pentru viitor. Așadar, s-au îndeplinit toate obiectivele enumerate la începutul stagiului doctoral.