

Rezumat al Tezei de Abilitare

Reducerea Decalajului dintre Știința Computațională a Rețelelor Complexă și Domeniul Ingineriei Calculatoarelor și Tehnologiei Informației

Dr.ing. Alexandru Topîrceanu

Știința rețelelor complexe este un domeniu interdisciplinar prin natură, cu rădăcinile sale încorporate în matematică, fizică, statistică, calculatoare și tehnologia informației. Cu toate acestea, aplicabilitatea domeniului se extinde, multidisciplinar, pentru a cuprinde științele biologice, farmacologice, sociale, economice, politice etc. Astăzi, cercetarea științei rețelelor (Network Science) poate fi clasificată în următoarele subiecte fundamentale: analiza rețelelor sociale (SNA), analiza computațională a rețelelor (CNA), analiza rețelelor biologice, rețele multistrat, analiza dinamică a rețelelor, analiza legăturilor (predicție a dinamicii rețelelor) și analiza de centralitate (influență în rețele).

Creșterea rapidă a științei rețelelor, în ultimele două decenii, inclusiv multe dintre metodologiile de modelare actuale - care implică cantități mari de date și sisteme complexe - sunt susținute de ingineria calculatoarelor și tehnologia informației. Ca atare, tehnologia informației oferă abordări în care algoritmii, instrumentele de simulare și bazele de date sunt utilizate pentru prelucrarea și înțelegerea datelor biologice, medicale, farmacologice sau sociale. Dezvoltările recente în domeniul educației și medicinei personalizate se bazează pe analiza și vizualizarea în Big Data și pe tehnologiile informatiche, cum ar fi analiza rețelelor complexe și învățarea automată. De asemenea, progresele în tehnologiile rețelelor sociale online au oferit oportunități în fizica socială și epidemiologia computațională cu impact socio-economic la scară globală.

Motivația principală a cercetării prezentate în această teză, cuprinsă între anii 2011-2021, este de a lega domeniul principal al calculatoarelor și tehnologiei informației cu domeniul multidisciplinar al științei rețelelor prin aplicabilitate reală, de impact. Această fuziune are ca rezultat termenul recent introdus de *Computational Network Science*, motivat de faptul că domeniul Calculatoare poate schimba toate celelalte științe prin abordarea sa bazată pe date. În sprijinul acestei tendințe recente, găsim noi reviste dedicate domeniului Știința Computațională a Rețelelor Complexă: grupul Nature a lansat o nouă revistă numită *Nature Computational Science* în 2021, încurajând dezvoltarea și utilizarea procedurilor de calcul și a modelelor matematice, inclusiv utilizarea acestora pentru a rezolva probleme complexe în diferite discipline științifice. De asemenea, IEEE a lansat revista *IEEE Transactions on Computational Social Systems* care abordează modelarea, analiza, simularea și înțelegerea sistemelor sociale dintr-o perspectivă cantitativă sau de calcul. În plus, este demn de remarcat faptul că printre personalitățile importante, considerate printre primii 10 cei mai puternici

oameni de știință din lume, găsim mai ales ingineri de calculatoare din mediul academic (de exemplu, Stanford, MIT) și din industrie (de exemplu, Google).

Având în vedere impactul științific ridicat și noutatea acestui domeniu, am început cercetarea în domeniul științei rețelelor în Universitatea Politehnica Timișoara încă din 2011. În timpul studiilor mele de doctorat (2012-2016) am publicat primele rezultate științifice folosind o nouă abordare computațională de rețea culminând cu doctoratul meu obținut în domeniul Calculatoare. Astfel, începând din 2016, am pus bazele ale unei noi Școli de Științe a Rețelelor Complexe în contextul Politehnic și, posibil, întregul context academic românesc.

Cercetările noastre s-au extins de-a lungul mai multor direcții, cum ar fi analiza rețelelor sociale (SNA), analiza computațională a rețelelor (CNA), medicina de rețea (Network Medicine). O primă contribuție remarcabilă a fost studierea generarea topologilor de rețea complexe realiste, evoluția lor în timp, urmată de difuziunea opiniei pe rețele mari aflate în evoluție. Pentru aceasta, analizăm modele topologice state-of-the-art, inspirate de modelele small-world și scale-free, cu scopul de a reproduce proprietățile observate empiric ale rețelelor din lumea reală. În acest sens, creăm un model de rețea socială foarte realist, utilizând o abordare algoritmică genetică multi-variabilă. De asemenea, spre deosebire de principiul fundamental al atașamentului preferențial bazat de gradul nodurilor (degree preferential attachment), susținut de A-L. Barabasi, am propus în continuare conceptul de atașament preferențial bazat pe Betweenness pentru a explica mai bine creșterea și dinamica rețelelor sociale. Ulterior, am studiat modele de răspândire a opiniei folosind opinii discrete și continue, inclusiv cele care includ agenți de tip stubborn. Având în vedere interacțiunea complexă a nodurilor (agenți), am ales simularea pe calculator ca metodologie validă de cercetare pentru a evalua și cuantifica aceste modele de răspândire a opiniei pe rețele sociale mari, deoarece o abordare analitică nu este capabilă să gestioneze complexitatea ridicată a inter-conectivității rețelelor sociale. Atunci când aplicăm puterea analitică a modelării și simulării computerizate, arătăm că noul nostru model bazat pe „toleranță”, pentru interacțiunea și răspândirea opiniei, generează modele realiste, reproductibile în rețelele sociale. În timp ce mai multe modele existente utilizează parametri de „încredere” pentru interacțiunea cu agentul, definim „toleranță” ca parametru de variabilă în timp și dependent de tiparele de interacțiune cu nodurile învecinate. Inspirația noastră pentru evoluția toleranței derivă din ideea că dinamica către toleranță și intoleranță variază exponențial, ceea ce înseamnă că un agent aflat sub influență constantă este indoctrinat cu o rată crescută în timp. De-a lungul acestei piste, am introdus și un instrument statistic pentru măsurarea similarității structurale între oricare două rețele complexe (fidelitatea statistică); am modelat antifragilitatea rețelelor complexe sub atac susținut; și, am introdus o metodologie originală și fiabilă pentru evaluarea comparativă a măsurilor de centralitate, într-un graf, într-un context competitiv. Există peste 50 de centralități diferite de noduri utilizate pentru selectarea nodurilor influente în rețele, iar metodologia noastră le poate selecta în mod fiabil pe cele mai eficiente având în vedere o topologie specifică.

O altă direcție importantă de cercetare detaliată în această teză este aplicarea medicinei de rețea (Precision Medicine, Systems Medicine) care oferă soluții computerizate pentru provocări medicale și farmacologice. Pe de o parte, am lucrat în ultimii opt ani în cercetarea

somnologiei, pentru a oferi soluții de calcul pentru prezicerea severității și dezvoltării apneei obstructive în somn (OSA) și a bolii pulmonare obstructive cronice (BPOC). Rezultatele se formează în jurul unui model de fenotip al pacienților, construit printr-o tehnică de grupare duală (dual clustering), rezultând un scor utilizabil de către medici în monitorizarea zilnică a pacientului. Mai recent, am extins state-of-the-art-ul în monitorizarea severității OSA, prezentând un model de fenotipare diferențiat pentru fiecare sex, precum și analizând cauzele răspunsului îmbunătățit la tratamentul CPAP. Toate aceste rezultate vizează dezvoltarea unui tratament personalizat și a unor diagnostice precise (cum ar fi sănătatea 4P). Pe de altă parte, folosim baza de date publică Drugbank pentru a construi o rețea de interacțiune medicament-medicament (DDI), în care nodurile reprezintă medicamente, iar legăturile reprezintă relații de interacțiune dintre medicamente. Mai recent, am explorat potențialul DDI bazat pe ținte farmacologice, folosind aceeași tehnică dual clustering. Aceste rezultate îi ajută pe cercetători să estimeze posibilele noi interacțiuni și să reutilizeze alternative pentru medicamente (repurposing), optimizând astfel studiile farmacologice costisitoare și consumatoare de timp.

O altă direcție întreprinsă de cercetare interdisciplinară a fost aceea de a adapta o abordare bazată pe rețele complexe în analiza datelor educaționale. Mai exact, am analizat datele studenților români participanți în cursuri MOOC, și am creat o rețea de compatibilitate a acestor studenți, pe baza motivațiilor, aşteptărilor și a dificultăților percepute de-a lungul cursurilor. Prin aplicarea tehnicilor de grupare în rețea, am definit arhetipuri specifice studenților. Mai mult, am dezvoltat o platformă de gamificare complet originală pentru motivația studenților, și am realizat un studiu asupra colaborării în sala de examen.

Rezultatele noastre științifice și academice sunt rezumate prin gestionarea a 2 proiecte naționale de cercetare (finanțate de UEFISCDI), participarea în 2 proiecte internaționale suplimentare (finanțate de Linde și Horizon 2020) și 5 proiecte naționale (finanțate de UEFISCDI și ARUT), publicarea a 3 cărți, peste 50 de lucrări indexate Web of Science, din care 16 reviste (12 indexate în Q1 / Q2), un factor de impact cumulat peste 45, un WoS h-index de 9 și 171 de citări (330 de citări în Google Scholar), calitatea de recenzent în diverse reviste multidisciplinare și IEEE, organizarea celor 9 ediții (2013-2021) ale competiției studențești SCMUPT pentru dezvoltarea de aplicații mobile, coordonare ca membru într-un comitet de doctorat, membru al mai multor comitete de raportare doctorală, și coordonarea a peste 90 de lucrări de licență și disertație.

Infrastructura de calcul disponibilă în prezent, așa cum a fost publicată pe platforma ERRIS, sprijină cercetarea privind modelarea și simularea, știința rețelelor, Big Data, dezvoltarea algoritmilor de grafuri și exploatarea datelor, toate acestea contribuind direct la domeniile de cercetare ale grupului nostru ACSA. În plus, infrastructurile oferite de laboratorul ACSA, platforma Vision NextCloud a departamentului, campusul virtual al universității și viitoarea platformă de calcul CloudPUTIng, disponibilă în viitor, vor oferi doctoranzilor un sprijin mai mult decât suficient pentru o carieră diversă de predare și cercetare.

Cu experiența noastră în coordonarea doctoranzilor pe o arie largă de subiecte științifice, considerăm grupul nostru de cercetare ACSA (Advanced Computing Systems and Architectures) ca o oportunitate atractivă pentru programele de doctorat. Cu cea mai nouă

includere a domeniului de Știința Rețelelor Complexe în portofoliul nostru de cercetare, suntem capabili să oferim un program de doctorat complet dezvoltat, axat pe Știința Rețelelor în domeniul Calculatoarelor și Tehnologiei Informației.

Eforturile noastre de cercetare se vor concentra pe două piste de cercetare mai recente (începând din 2018), cea a epidemiologiei computaționale și a predicției sondajelor politice. Având în vedere pandemia COVID-19, am fost rapid motivați să colaborăm cu Universitatea din Texas și să ne concentrăm eforturile asupra modelării și înțelegerei mai bune a impactului acestei pandemii. Cercetările actuale sunt în curs de dezvoltare pe două căi: aceea de a înțelege impactul strategiilor de izolare adoptate la începutul anului 2020 și de a îmbunătăți modelarea populației și a mobilității indivizilor utilizate pentru simulările epidemiologice. De asemenea, având în vedere frământările provocate de alegeri locale și globale din ultimii ani, ne-am concentrat pe îmbunătățirea preciziei sondajelor pre-electorale folosind analiza seriilor de timp și știința rețelei. Am obținut rezultate încurajatoare în comparație cu metodele existente de ultimă generație, cum ar fi regresia pe mai multe niveluri cu post-stratificare.

În paralel, ne propunem să dezvoltăm noi propuneri de proiecte, cum ar fi un proiect TE (Tinere Echipe) și PED (Proiect Experimental Demonstrativ). Având în vedere actualii membri ACSA, găsim ambele alternative fezabile. Un proiect TE poate fi țintit în domeniul epidemiologiei computaționale (un sub-domeniu actual, de impact, al științei computaționale a rețelelor). Un PED poate fi vizat de implementarea unui instrument legat de cercetarea somnului (pentru OSA). De asemenea, proiectul început Horizon 2020 (2021-2025), în care sunt implicați mai mulți membri ai echipei ACSA și UMFT, deschide posibilități multiple pentru masteranzii și pentru doctoranzii actuali și viitori. Având în vedere numărul mare de centre de cercetare implicate în proiect, majoritatea publicațiilor vor avea un impact științific solid, aceasta deschizând oportunități de vizibilitate pentru toți participanții.

Toate realizările științifice ale candidatului tezei ilustrează potențialul ridicat de cercetător independent și capacitatea sa de a gestiona echipe de cercetare și viitorii doctoranți. El colaborează activ cu mai multe grupuri de cercetare, cu unele rezultate notabile: Universitatea Carnegie Mellon / Universitatea din Texas, Universitatea Southern California (USC) și Universitatea Central Europeană (vizită de cercetare susținută de proiectul IMPRESS). În viitorul apropiat propunem stabilirea unor colaborări cu grupul DSG de la Universitatea de Știință și Tehnologie din Wrocław, grupul Collide de la Universitatea Duisburg-Essen și departamentul DNDS de la CEU. Aceste colaborări științifice ar trebui să consolideze vizibilitatea internațională, originalitatea ridicată a cercetării și relevanța activității grupului ACSA, deoarece toate echipele menționate lucrează în domeniul științei rețelelor complexe.

Această teză servește drept o dovedă puternică a cercetării cu impact ridicat care poate fi realizată prin angajarea științei și ingineriei calculatoarelor în domenii multidisciplinare. Ca atare, intenționăm să restrângem diferența dintre calculatoare și știința rețelelor prin abordarea în continuare a unor subiecte de cercetare provocatoare din diverse domenii ale științei, prin participarea la principalele conferințe din domeniul nostru, prin stabilirea unei colaborări internaționale de durată, prin crearea de parteneriate de proiect, prin inițierea a noi cursuri specializate de licență în departamentul nostru, și prin integrarea de noi doctoranți în grupul nostru de cercetare ACSA.