Rezumatul tezei de abilitare

Conf.dr.ing. Guillaume DUCOFFE

în vederea obţinerii atestatului de abilitare

Pe parcursul acestui document, prezint o parte din rezultatele ştiintifice cu impact pe care le-am obţinut de când am susţinut teza mea de doctorat (Dec. 2016). Problemele principale pe care le consider în această abilitare sunt calcularea ***diametrului*** şi respectiv a ***unui cuplaj maxim*** ale unui graf, ambele având aplicaţii în Analiza Reţelelor. Aceste probleme fundamentale pot fi rezolvate în timp polinomial. Totodată, existenţa unor algoritmi în timp cvasi-liniar a rămas o problemă larg deschisă. În cele ce urmează, prezint mai multe căi pentru a îmbunătăţi timpii de rulare ale acestor probleme, dar asumând că grafurile de intrare sunt restricţionate la unele clase cu mai multă structură. O astfel de ipoteză, restricţionând tipurile de intrare, are sens în contextul studiului reţelelor reale şi dezvoltării unor reţele tehnologice.

Prima parte din acest document este dedicată la unele probleme despre cuplaje întrun graf. Existenţa unui algoritm în timp cvasi-liniar pentru calcularea unui cuplaj maxim este o problemă deschisă de foarte multă vreme. Se prezintă aici un astfel de algoritm în cazul în care grafurile de intrare au parametrul ***clique-width*** limitat la o constantă. Acest algoritm exploatează teorema lui Courcelle, ceea ce este interesant în sine deoarece nu se poate exprima un cuplaj folosind logica monadică $CMSO_1$ — această condiţie fiind cea standardă pentru a aplica teorema. Existenţa unui algoritm al cărui timp de rulare depinde ***polinomial*** de parametrul clique-width a rămas o problemă deschisă. Sunt prezentaţi şi câţiva algoritmi parametrizati pentru calcularea unui cuplaj maxim în timp cvasi-liniar, care depind polinomial de unii parametrii mai slabi decât clique-width, precum ***modular-width*** şi P_4-sparseness.

În a doua parte a documentului, se consideră problema diametrului şi alte probleme legate cu distanţe întrun graf, pentru care nu se crede că ele pot fi rezolvate în timp sub-cvadratic. Rezultatele din această parte sunt motivate de conjectura mea informală că toate aceste probleme pot fi rezolvate în timp sub-cvadratic dacă grafurile de intrare îndeplinesc o proprietate asemănătoare cu cea a lui Helly. Dintre aceste rezultate, se poate menţiona în special un algoritm pentru calcularea în timp sub-cvadratic a excentricităţilor unui graf care exclude un graf H oricare ca minor. Anterior, acest rezultat era cunoscut numai pentru grafurile planare (Cabello, *SODA'16*) şi grafurile ale căror parametrul treewidth este limitat la o constantă (Cabello & Knauer, *Computational Geometry*, 2009). Algoritmul este bazat pe ***teoria Vapnik-Chervonenkis***. Din această teorie, se pot deduce şi unele rezultate noi despre grafurile cordale şi despre alte clase de grafuri. Totodată, şi ***grafurile Helly*** sunt foarte mult studiate, datorită proprietăţii că

orice graf este subgraful izometric al unui graf Helly. Sunt introduse unele tehnici noi pentru rezolvarea în timp sub-cvadratic a unor probleme legate cu distanţe în grafurile Helly. Abordarea propusă aduce la o generalizare semnificativă a unor rezultate obţinute anterior pentru grafurile interval (Olariu, *International Journal of Computer Mathematics*, 1990) şi grafurile dual cordal (Brandstädt et al., *Discrete Applied Mathematics*, 1998).

Cuvinte Cheie: Diametrul unui graf; Cuplaj; Clique-width; dimensiunea Vapnik-Chervonenkis; proprietatea lui Helly; Algoritmica grafurilor; Teoria complexităţii.

Se raportează în acest document activităţile mele de cercetare şi de predare de când am susţinut teza mea de doctorat (Dec. 2016). Lucrarea mea de cercetare este motivată de unele aspecte computaţionale ale Analizei Reţelelor. În timpul desfăşurării tezei mele de doctorat, am studiat proprietăţile metrice ale unui graf (din care, în special, hiperbolicitatea Gromov) pentru a explica mai bine performanţele practice ale unor protocoale de rutare ale reţelelor complexe. În cercetarea mea cea mai recentă, am adresat limitele computaţionale ale metodelor de rezolvare ale unor sarcini specifice reţelelor (abstractizate în unele probleme ale grafurilor) pentru sistemele complexe de mare dimensiune. Aceste rezultate au fost obţinute în cadrul poziţiilor didactice şi de cercetare pe care le-am ocupat de când am ajuns în Romania (Ian. 2017), la institutele de cercetare ICUB şi ICI Bucureşti, şi la Facultatea de Matematică şi Informatică a Universităţii din Bucureşti.

Activităţile mele de cercetare, începând cu anul 2017 până în momentul de faţă, pot fi împărţite în trei perioade suprapuse. În linii mari, aceste perioade mai sus menţionate se raportează la temele mele de cercetare principale din ultimii ani. În cele ce urmează, se referă şi la capitolele şi secţiunile în care mai multe informaţii despre fiecare temă pot fi găsite.

• Perioada 2017–2018: bursa postdoctorala ICUB, având ca obiect principal studiul puterilor grafurilor (Secţiunea A.3 din anexa).

• Perioada 2018–2021: studiul despre complexitatea parametrizată a problemei calculării unui cuplaj maxim întrun graf (Partea I). Rezultatele din această parte au fost obţinute în colaborare cu D. Coudert si A. Popa.

• Perioada 2020–2023: îmbunătăţirea complexităţii calculării diametrului unui graf, în cazul unor clase de grafuri bine structurate (Partea II). Rezultatele din această parte au fost obţinute în colaborare cu F. Dragan, M. Habib, H. Guarnera şi L. Viennot.

Conţinutul tehnic al acestui document (Capitolele 3 – 10) are la bază ultimele două perioade menţionate mai sus. Totodată, o prezentare scurtă a proiectului meu de cercetare postdoctoral poate fi regasită în Anexa A.

În cele ce urmează, întâi voi raporta activităţile mele de predare (Secţiunea B.1). În special, în Secţiunea B.1.1, prezint pe scurt experienţa mea de îndrumare a studenţilor. Principalul motiv pentru care am scris această teza de abilitare este de a obţine dreptul de a coordona doctoranzi, pe temele de cercetare care au fost iniţiate în publicările şi proiectele mele. Restul acestei lucrări este dedicat excluziv activităţilor mele de cercetare.

Contribuţiile pe care le prezint în această lucrare sunt de natură teoretică. Aceste rezultate se încadrează în domeniul Teoriei Grafurilor, care este introdus pe scurt în Secţiunea B.2. Apoi, în Secţiunea B.3, se pun în context obiectivele mele de cercetare. Se prezintă în mod general întrebările deschise principale din domeniu, şi abordările mele pentru a le rezolva. Secţiunea B.4 introduce două probleme algoritmice de prim plan, precum şi contribuţiile mele corespunzătoare, care sunt detaliate în capitolele tehnice ale acestui document. Aceste două probleme sunt calcularea diametrului şi a unui cuplaj maxim întrun graf. Totodată, menţionez şi că în Anexa A, se pot regăsi mai multe contribuţii importante despre problemele Leaf Power şi calcularea hiperbolicităţii Gromov. Secţiunea B.5 este un sumar detaliat al capitolelor tehnice.

B.1 Activităţi de Predare

Începând cu anul 2017, am fost implicat în programul de master „Ingineria Software", din cadrul Facultăţii de Matematică şi Informatică a Universităţii din Bucureşti. Am ţinut mai multe lecţii, în legatura (directă şi/sau indirectă) cu activităţile mele de cercetare. Primul curs pe care l-am ţinut în acest program de master: „Evaluarea Performanţelor Sistemelor Complexe", era bazat pe teoria lanţurilor markoviene. Această teorie a stat la baza studiilor mele despre grafurile aleatoare (considerate ca fiind modele ideale pentru reţelele complexe), în perioada tezei mele şi după aceasta. În prezent (începând cu anul 2022), ţin un curs despre algoritmi paraleli şi distribuiţi pentru grafuri, ceea ce consider sa fie o tematică pertinentă în contextul analizelor la scara largă ale unor mulţimi de date relaţionale. De asemenea, am propus şi un curs opţional despre reţelele complexe, în contextul căruia se discută despre caracteristicile acestor reţele, şi aplicaţiile algoritmice ale acestora.

Începând cu anul 2018, o dată ce am obţinut titlul academic de lector (am caştigat titlul de conferenţiar, prin concurs, în anul 2021), am predat şi cursuri de licenţă. Mai exact, am ţinut orele de Teoria Grafurilor (până în anul 2020), şi de Structuri de Date şi Algoritmică. Ţinând cont de cunoştinte acumulate de către studenţii noştri la liceu, am avut ocazia să predau şi câteva subiecte mai avansate, precum: metodele eficiente de a determina cei mai mici stramoşi comuni întrun arbore cu radacina, structura de date range tree, cuckoo hashing, algoritmul de sortare Timsort, etc. Începând cu anul 2022, am propus un curs opţional despre algoritmica avansată a grafurilor, al cărui conţinut este asămănator cu cel al mai multor cursuri din străinătate, precum cele din Masterul Parizian de Cercetare în Informatică (MPRI).

Totodată, în perioada 2019-2021, am fost invitat să ţin şi cursurile de Algoritmică şi Structuri de Date la Universitatea Tehnică de Construcţii Bucureşti.

B.1.1 Îndrumarea Studenţilor

Am îndrumat în jur de şapte lucrări de licenţă (cinci lucrări de când am ajuns în România), şi trei lucrări de master (două lucrări de când am ajuns în România). În general, accept să coordonez doar lucrările ale căror temă are o legătură cu activităţile mele de cercetare, ceea ce explică numărul relativ redus de lucrări coordonate. Am găzduit şi T. Trolliet, în timpul tezei sale de doctorat, pentru o vizită de cercetare. Am fost parte din două comisii de îndrumare a unor doctoranzi.

În perioada 2022-2023, am implicat doi studenţi de la facultate întrun Proiect Experimental Demonstrativ (PED) pentru care am fost director de proiect. În viitor, intenţionez să includ şi studenţii mei doctoranzi în proiectele mele în derulare.

B.2 Domeniul Studiat

Pe parcursul acestui document, o reţea este întotdeauna reprezentată ca fiind un graf. Reamintim ca un graf este definit ca fiind o pereche G = (V, E) astfel încât V se numeşte mulţimea vârfurilor (”nodurile reţelei”) şi $E \subseteq V^2$ se numeşte mulţimea muchiilor (”link-urile reţelei”). Cititorii acestui document se pot referi la Capitolul 2 pentru mai multe definiţii şi terminologii.

În modelarea unor cazuri reale, nodurile şi link-urile unei reţele pot fi înterpretate în mai multe feluri. Această flexibilitate ne permite să folosim modelele bazate pe grafuri în studiul mai multor sisteme complexe (fie biologice sau tehnologice): ceea ce include reţelele neuronale, reţelele sociale, reţelele de interacţiuni dintre proteine, şi centrele de date. Totodată, dacă acest lucru se potriveşte cu sistemul studiat, este posibil şi să orientăm muchiile, să alocăm ponderi şi/sau etichete la fiecare nod/muchie, sau să înlocuim muchiile cu unele submulţimi ale nodurilor cu dimensiunea nenulă oricare. Atunci ne referim fie la un graf orientat, la un graf ponderat, sau la un hipergraf, şi la terminologia corespunzătoare. Aceste modele bazate pe grafuri augmentate pot aduce la o îmbunătăţire a modelării unor cazuri reale. De exemplu, existenţa unor restricţii tehnologice, cum ar fi interferenţele radio sau latenţa unui link, poate fi integrată în model prin alocarea unor etichete sau a unor ponderi. Totodată, modelul standard în Analiza Reţelelor rămâne cel bazat pe grafuri neorientate şi neponderate, acesta fiind un compromis satisfăcător dintre simplitate şi realism. Mai mult, este de înţeles că rezolvarea unor probleme ale sistemelor complexe cere întâi rezolvarea acestor probleme pentru grafurile de bază. *Dacă nu se specifică altfel, rezultatele prezentate în această teza de abilitare sunt numai pentru grafurile neorientate şi neponderate*.

Teoria Grafurilor se împarte în mai multe subteorii (ex., structurală, metrică, algebrică, topologică, etc.), cu unele abordări, implicaţii şi rezultate diferite. În cadrul *Algoritmicii Grafurilor*, un graf este o structură de date abstractă. Problemele computaţionale ale grafurilor sunt mereu motivate de studiul unor proprietăţi matematice ale acestora, sau de o aplicaţie practică. Întradevar, sunt unele sarcini ale reţelelor, precum alocarea resurselor şi transferul de pachete, care pot fi uşor modelate ca unele probleme ale grafurilor. În mai multe cazuri, obiectivul unei astfel de probleme constă în calcularea mai multor şiruri de muchii, având proprietatea că fiecare două muchii consecutive întrun şir trebuie fi incidente, şi cu anumite restricţii suplimentare care variază între problemele. În cele ce urmează, un şir de muchii cu proprietatea, mai sus menţionata, că fiecare două muchii consecutive sunt incidente, se numeşte abuziv un *lanţ* (muchiile şi ciclurile sunt considerate şi ele ca fiind lanţuri). O teorie a problemelor bazate pe lanţuri a fost iniţiată de către Tarjan. De exemplu, problemele de calculare a unui flux, a unui cuplaj, sau a unor drumuri minime întrun graf, sunt toate bazate pe lanţuri. Aceste probleme se potrivesc în special cu studiul reţelelor de comunicaţii (ex. reţelele de telecomunicaţii, reţelele sociale, reţelele de transport, reţelele de interacţiuni, etc.). *Toate problemele considerate în această lucrare sunt bazate pe lanţuri*.

B.3 Temele de cercetare (Prezentare informală)

În linii mari, activităţile mele de cercetare sunt motivate de următoarea întrebare informală: *Să se determine structurile subiacente întrun graf care pot fi folosite pentru a rezolva cât mai eficient unele probleme „dificile" care sunt bazate pe lanţuri*. Ceea ce înseamana că în acest caz eficienţa unui algoritm depinde de cum definim o problemă dificilă. Dacă este vorba de o problemă NP-hard, atunci obiectivul ar fi să identificăm unele clase de grafuri cât mai generale astfel încât problema să poată fi rezolvată în timp polinomial[1]. Această abordare este conform tezei lui Cobahm care caracterizează algoritmii „eficienţi" ca fiind cei care rulează în timp polinomial. Totodată, s-a dezvoltat recent şi o teorie a dificultăţii problemelor din clasa P. Nevoia pentru o astfel de teorie mai fină este parţial motivată de contextul tehnologic. Întradevăr, trecerea noastră foarte rapid de la Revoluţia Industrială până la Era Informaţională a rezultat întro supradimensionare fără precedent a reţelelor de comunicaţii, ceea ce s-a văzut în special cu Internet-ul. La momentul redactării tezei mele de doctorat, se aşteptase ca volumul traficului Internet global să fi ajuns la 40 zettabyte până în anul 2020. Dar această predicţie a fost depaşită cu mult, volumul real în anul 2020 fiind de 64.2 zettabyte, şi se asteaptă acum ca aceasta să depăşească 180 zettabyte până în anul 2025. În acest context, ar fi mai rezonabil să definim o problemă ca fiind dificilă dacă aceasta nu poate fi rezolvată în timp $O(N^{q-\varepsilon})$, N fiind dimensiunea intrării, pentru o anumită constantă q destul de mică (să zicem de exemplu $q \leq 3$) şi orice valoare $\varepsilon > 0$. Pentru mai multe detalii, cititorii interesaţi pot consulta Capitolul 2. Cele două probleme principale despre care se discută în această lucrare: problemele

[1] Determinarea existenţei unui algoritm în timp *cvasi-polinomial*, adică, care rulează în timp $O(2^{\text{polylog}(N)})$, N fiind dimensiunea intrării, poate fi şi ea o provocare interesantă. Întradevăr, sub unele ipoteze de complexitate standarde, precum Ipoteza Timpului Exponenţial (ETH), o problemă NP-hard nu poate fi rezolvată pentru grafurile oarecare nici în timp cvasi-polinomial.

calculării *diametrului* şi a unui *cuplaj maxim* întrun graf, sunt două exemple de probleme dificile din clasa P.

Merită menţionat că cercetarea mea *nu* are ca scop dezvoltarea unor algoritmi eficienţi în practică. În linii mari, dându-se unele probleme care pot fi rezolvate în timp $O(N^q)$, pentru o anumită constantă $q > 1$, studiez proprietăţile adiţionale suficiente pentru existenţa unui algoritm în timp $O(N^{q-\varepsilon})$, pentru o valoare $\varepsilon > 0$ oricare. Totodată, o astfel de îmbunătăţire a exponentului este pertinentă doar în cadrul studiului asimptotic al algoritmului. În practică, alte aspecte precum constanta ascunsă în notaţia ”O-Mare” (O) şi folosirea unor structuri de date cât mai simple se pot revela mult mai importante. Implementarea şi testarea în practică a unor algoritmi este un domeniu de cercetare separat, care se numeşte Ingineria Algoritmilor (*Algorithm Engineering*). Rezultatele mele de cercetare pot fi considerate ca o dovadă de fezabilitate pentru existenţa unui algoritm practic.

Un algoritm presupunând că grafurile de intrare îndeplinesc anumite restricţii netestate, poate să nu termine pentru un graf oricare (dacă acesta nu îndeplineşte restricţiile menţionate mai sus) sau să returneze o ieşire greşită. Din fericire, se ştie că grafurile reale *nu* sunt oarecare. Există o amplă literatură despre proprietăţile acestora, precum fenomenul lumea (foarte) mică – *(ultra) small-world phenomenon* –, clustering, şi distribuţia gradelor după regula puterilor. A se vedea teza de doctorat a lui T. Trolliet pentru mai multe detalii şi exemple. De asemenea, se pot folosi unele metode de construire specifice, astfel încât reţele tehnologice (ex., reţelele integrate în sateliţi) să indeplinească proprietăţile necesare.

Dându-se o problemă de rezolvat întro clasă de grafuri C, se defineşte un *algoritm robust* ca fiind un algoritm A care pentru orice graful G de intrare, fie rezolvă corect problema, fie certifică că $G \notin C$. Această definiţie a fost introdusă de către Spinrad. Se poate observa că prin definiţie, un algoritm robust nu poate returna o ieşire gresită (deşi acesta tot poate să nu termine pentru grafurile oarecare care nu sunt în clasa C). În contrast, nu se poate avea încredere în ieşirea unui algoritm ne-robust. Deşi orice algoritm pentru grafurile întro clasă C poate fi transformat întrun algoritm robust, prin integrarea unui test preliminar dacă graful de intrare aparţine clasei, o astfel de operaţie poate impacta dramatic timpul de rulare al algoritmului. De exemplu, am prezentat în [Duc23] un algoritm în timp $\tilde{O}(m\sqrt{(kn)})$[2] pentru a calcula raza unui graf cu n varfuri si m muchii al carui număr Helly este cel mult egal cu k (a se vedea şi Capitolul 8). Totodată, problema calculării numărului Helly este coNP-hard[3]. Pe de o parte, folosirea unor algoritmi ne-robuşti pentru o clasă de grafuri C trebuie să fie limitată la subclasele $C' \subseteq C$ astfel încât se

[2] Se foloseşte notaţia Õ pentru a ignora dependinţele poli-logaritmice în timpul de rulare.

[3] Acest rezultat a fost demonstrat doar pentru hipergrafuri, dar aceeaşi abordare poate fi aplicată şi la grafurile split.

poate verifica întrun mod eficient dacă un graf este în C′. Pe de o altă parte, un algoritm pentru grafurile din C poate fi aplicat la grafurile din orice subclasă C′.

În cazul reţelelor complexe, caracteristicile specifice ale acestora pot fi deduse de o analiză a unor mulţimi de date reale. Atunci, în timpul implementării unui algoritm nou, ne putem restricţiona la grafurile cu unele proprietăţi asemănătoare reţelelor complexe. O abordare alternativă constă în dezvoltarea unor modele de grafuri aleatoare, ceea ce putem considera ca fiind modele ideale pentru reţelele complexe. Dacă aceste grafuri aleatoare îndeplinesc o anumită proprietate P cu mare probabilitate, şi mai mult există o corelaţie pozitivă dintre modelul corespunzător şi datele reale, atunci un algoritm poate fi folosit ca euristica dacă acesta este demonstrat ca fiind corect pentru grafurile cu proprietatea P. În acest document, nu se detaliează mai multe legate de acestea aspecte.

Exemple. Teoria locaţiei este strâns legată cu activităţile mele de cercetare recente. Pentru a ilustra această relaţie, să se considere problema amplasării unor job-uri întrun centru de date. Ca să se poată accesa şi împărţi resursele reţelei, este nevoie de mai multe transferuri costisitoare ale căror costul poate reprezinta până la jumătatea timpului de finalizare. Întro astfel de situaţie, este de dorit ca fiecare job să fie programat astfel încât număr de transferuri să fie minimizat. – Bineînţeles, acesta este o simplificare. De fapt, sunt mai multe criterii de considerat. – Acest obiectiv poate fi re-formulat ca fiind o problemă a grafurilor. Mai specific, dându-se o submulţime S de vârfuri (locaţiile resurselor), un vârf se numeşte *central* dacă acesta minimizează distanţă maximă până la un vârf din S; se numeşte *mediană* un vârf care minimizează distanţa medie până la un vârf din S. Problema *centrului* cere să se calculeze un vârf central întrun graf; de asemenea, problema *medianei* cere să se calculeze o mediană întrun graf[4]. Pentru un alt exemplu, în domeniul calculării paralele, să se considere difuzarea unor mesaje dintre surse şi destinaţii oarecare. Atunci, *diametrul* (distanţa maximă dintre două vârfuri) este timpul maxim de finalizare a trimiterii, în lipsa unor interferenţe. Problema *diametrului* cere să se calculeze diametrul unui graf. A se vedea Partea II a acestei lucrării pentru un studiu al complexităţii acestor probleme.

Dacă pentru fiecare vârf v se execută o parcurgere în laţime (BFS), atunci se obţine un arbore al drumurilor minime cu rădăcina v. O astfel de parcurgere se face în timp liniar. Facând aşa, se poate calcula matricea distanţelor întrun graf în timp polinomial. În special, se pot rezolva în acelaşi timp problemele centrului, ale medianei şi ale diametrului. Totodată, aceşti algoritmi ”naivi” nu sunt practici. Întradevăr, timpul lor de rulare este cvadratic în numărul de muchii[5]. Nu este surprinzător că se consideră

[4] În general, se presupune că S = V (G), adică, toate vârfurile sunt considerate în egală măsură. Totodată, s-a studiat şi cazul general cu S fiind oricare, sub denumirile problemei 1-Center şi problemei lui Weber.
[5] Este şi cubic în numărul de vârfuri. Totodată, având în vedere că, în general, numărul de muchii întro reţea complexă este proporţional liniar cu numărul de vârfuri, dependinţa în numărul de muchii este cea mai pertinentă.

o complexitate cvadratică ca fiind una excesivă pentru grafurile mari, cu milioane de vârfuri şi uneori miliarde de muchii. *Depăşirea barierei cvadratice* poate fi văzută ca fiind un prim pas spre dezvoltarea unor algoritmi practici.

Din păcate, există unele probleme care nu pot fi rezolvate în timp subcvadratic. Spre exemplu, este cazul de problema calculării distanţelor intrun graf, deoarece aceasta cere să se calculeze o matrice cu un număr de intrări pătratic. Totodată, dacă ieşirea problemei este limitată la o singură valoare, ceea ce este cazul pentru problema diametrului, atunci nu se poate obţine cu această abordare decât o margine inferioară trivială. Conform teoremei de ierarhie în timp, există unele probleme decizionale – aşadar, ale căror ieşire este un singur bit! – care pot fi rezolvate în timp cvadratic dar nu pot fi rezolvate în timp subcvadratic. Dar aceste probleme sunt artificiale, fiind obţinute printrun argument de diagonalizare. Acum zece ani, înca nu exista o teorie a complexităţii care ar fi permis să demonstrăm că o problemă nu poate fi rezolvată în timp subcvadratic, sub unele ipoteze standarde. Dar între timp a apărut o teorie modernă a dificultăţi problemelor din clasa P, bazată pe unele metode noi de reducere dintre problemele polinomiale. A se vedea Capitolul 2 pentru mai multe detalii. În special, a rezultat din această teorie o înţelegere cât mai bună despre complexitatea problemei diametrului, atât pentru calcularea exactă cât şi pentru calcularea aproximată. În linii mari, dificultatea acestei probleme a fost studiată prin cazul special de a decide dacă diametrul unui graf este egal cu doi. Putem observa că diametrul unui graf este cel puţin egal cu trei dacă şi numai dacă există două vârfuri ale căror vecinităţi sunt disjuncte. Această relaţie dintre problema diametrului şi problema mulţimilor disjuncte a fost studiată pentru prima dată în anul 1992, dar fără să atragă atenţia algoritmicienilor. Recent, Williams a demonstrat că problema mulţimilor disjuncte nu poate fi rezolată în timp subcvadratic sub ipoteza timpului exponenţial puternic (SETH). – Pe baza aceasta, s-a re-descoperit relaţia dintre problema diametrului şi problema mulţimilor disjuncte, ceea ce a adus la rezultatul de imposibilitate de a calcula diametrul unui graf în timp sub-cvadratic sub ipoteza timpului exponenţial puternic. – Ipoteza timpului exponenţial puternic este o versiune mult mai puternică de problemă P vs. NP. Mai specific, se face ipoteza că problema satisfiabilităţii nu poate fi rezolvată în timp $2^{\epsilon n}$, pentru orice $\epsilon < 1$. Este surprinzător cât de mult o conjectură despre problemele NP-hard a contribuit la înţelegerea noastră despre problemele din clasa P.

Abordarea mea în cercetare constă în exploatarea algoritmică a unor proprietăţi structurale şi metrice ale grafurilor, care sunt deseori întâlnite în reţele complexe. Spre exemplu, putem studia complexitatea calculării diametrului în cazurile următoare:

• *diametrul este proportional cu numărul de vârfuri.* — Studiul acestui caz este pertinent în analiza reţelelor rutiere şi a reţelelor moleculare, în care s-au observat existenţa a mai multor sub-structuri înlănţuite.

• *graful este planar, sau poate fi incoporat întro suprafaţă a cărei gen este limitat la o constantă, în aşa fel încât muchiile să se intersecteze doar în noduri.* — Studiul acestui caz poate avea aplicaţii în analiza reţelelor străzilor, a reţelelor rutiere, etc.

• *graful exclude un anumit pattern* (ex: subgrafuri induse, minori, mulţimi asteroidale, etc.). — Există o astfel de caracterizare pentru grafurile interval, şi toate clase de grafuri ereditare.

Alături de co-autorii mei, am rezolvat, complet sau parţial, toate cele trei cazuri menţionate mai sus. Merită menţionat că s-au dedus răspunsurile lor dintro singura *proprietate comună* a acestor grafuri. O astfel de generalitate este rareori întâlnită în literatura de specialitate.

Algoritmii mei sunt mereu completaţi cu unele demonstraţii de margini inferioare (condiţionate la unele ipoteze). Spre exemplu, am demonstrat că diametrul unui graf cu n vârfuri poate fi calculat în timp cvasi-liniar daca acesta este în $\omega(n/ \log n)$. Totodată, dacă ipoteza timpului exponenţial puternic (SETH) este adevarată, atunci pentru orice valoare $\varepsilon > 0$, diametrul unui graf cu n vârfuri nu poate fi calculat în timp $O(n^{2-\varepsilon})$, chiar dacă acesta este în $o(n/ \log n)$. A se vedea şi Capitolul 10. Aceste rezultate de mai sus contribuie la o mai bună înţelegere despre sursa dificultăţii a problemelor considerate.

B.4 Contribuţiile cele mai semnificative

În cele ce urmează, se discută numai despre rezultatele pe care le-am obţinut după susţinerea tezei mele de doctorat (începând de luna Ian. 2017). În special, nu am luat în considerare contribuţiile mele în domeniul *Web's Transparency* (ceea ce include şi articolul meu cel mai citat[6]), pentru care încă nu am găsit colaboratori noi de când am ajuns în România. De asemenea, nu am luat în considerare rezultatele pe care le-am obţinut, alături de îndrumătorul meu de doctorat D. Coudert, despre descompunerea unui graf prin *clici separatoare*. Aceste rezultate încă continuă să fie citate şi folosite deseori în unele articole ştiintifice din domeniile algoritmicii şi biologiei.

Rezultatele prezentate în această lucrare sunt structurate în jurul a două probleme fundamentale ale grafurilor: problemele calculării a unui *cuplaj maxim* şi a *diametrului* unui graf. În cele ce urmează, sunt introduse amandouă problemele, şi contribuţiile corespunzătoare.

B.4.1 Contribuţii pentru probleme despre cuplaje (Partea I)

Un *cuplaj* întrun graf este o mulţime de muchii astfel încât nu sunt două muchii cu un capăt comun. Problema cuplajului maxim cere să se calculeze un cuplaj al cărui număr de muchii este

[6] Merită menţionat că tema articolului ţine mai mult de domeniul Ştiinţelor Inginereşte decât de domeniul Informaticii Teoretice, ceea ce explică parţial numărul lui de citări mult mai ridicat.

maximizat. În special, un cuplaj este *perfect* dacă fiecare vârf este incident cu o muchie a cuplajului. Se poate observa uşor că orice cuplaj perfect este şi un cuplaj maxim, deşi un cuplaj perfect poate să nu existe. Problema calculării unui cuplaj maxim este o sarcină de bază în mai multe mecanisme de alocare a resurselor, în scheduling, dar şi în chimie, deoarece sa făcut ipoteza că existenţă unui cuplaj perfect (ceea ce se numeşte şi o structură Kekulé în acest context) este o condiţie necesară pentru un sistem conjugat să se sintetizeze. Cel mai bun algoritm cunoscut pentru această problemă rulează în timp $O(m\sqrt{n})$ pentru grafurile oarecare cu n vârfuri şi m muchii. Acesta este aşa de complex că mai avem nevoie de corecţi şi de clarificări peste două zeci de ani de când s-a publicat articolul original.

Unii autori au sugerat că este pertinent să luam în considerare problema cuplajului maxim în cadrul dezvoltării unei teorii a complexităţii parametrizate pentru problemele din clasa P. Sugestia lor a fost parţial motivată de existenţa unor metode de prelucrare pentru problema cuplajului maxim care sunt foarte eficiente în practică. Aceşti autori au demonstrat existenţa mai multor algoritmi parametrizaţi pentru această problemă, care rulează în timp $f(k)(n+m)^{1+o(1)}$, f fiind o funcţie a parametrului k considerat. Aceste rezultate, obţinute prin studierea mai multor parametrii, au dus la o unificare şi la o generalizare a mai multor algoritmi cunoscuţi anterior pentru unele clase de grafuri izolate. Într-adevăr, algoritmii parametrizaţi pot fi consideraţi ca o metodă de lucru pentru a unifica şi a generaliza rezultatele deja cunoscute despre unele clase de grafuri diferite, prin folosirea unor proprietăţi comune. Ceea ce mă interesează cel mai mult în activităţile mele de cercetare este de a identifica cele mai generale proprietăţile ale grafurilor care pot fi exploatate pentru a rezolva mai rapid o problemă. A se menţiona că cu cât mai generale sunt proprietăţile considerate, cu cât mai puţin putem îmbunătăţi timpul de rulare.

Parametrii *Tree-width* şi *Clique-width* sunt printre cei mai studiaţi în literatura. A se vedea Capitolul 2 pentru definiţia lor. Există unele relaţii între aceşti parametri şi unele moduri de reprezentare ale unui graf ca un arbore, ceea ce permite dezvoltarea mai multor algoritmi pentru aceste grafuri, folosind programarea dinamică de pe reprezentarea lor arborescentă. Aşadar, sunt mai multe probleme NP-hard care pot fi rezolvate în timp polinomial pentru grafurile ale căror parametru treewidth sau clique-width este limitat la o constantă. Au fost prezentaţi unii algoritmi parametrizaţi pentru calcularea unui cuplaj maxim întrun graf, care rulează în timp cvasi-liniar şi depind polinomial de treewidth. Ceea ce ne a motivat pe mine, D. Coudert şi A. Popa, să studiem existenţa unui astfel de algoritm parametrizat care tot rulează în timp cvasi-liniar dar depinde polinomial de clique-width (acesta fiind un parametru mai general decât tree-width).

Rezultate. *Lanţurile de creştere* stau la baza mai multor algoritmi pentru calcularea unui cuplaj maxim, existenţa unora fiind, în linii mari, un certificat care dovedeşte că un cuplaj nu este maxim. Am prezentat unele relaţii structurale dintre lanţurile de creştere şi *modulele* întrun graf. Un modul al unui graf G = (V,

E) constă întro submulţime M de vârfuri în aşa fel încât vecinii lor în V \M sunt identici. *Descompunerea modulară* a unui graf este o descompunere recursivă în module ale acestui graf. Există unele relaţii importante între modulele unui graf şi parametrul clique-width. A se vedea şi Capitolul 2. Totodată, s-au observat şi unele relaţii dintre modulele unui graf şi structura comunităţiilor întro reţea complexă. Cu ajutorul unor rezultate structurale noi, am obţinut unii algoritmi în timp liniar pentru calcularea unui cuplaj maxim în mai multe subclase de grafuri ale căror parametru clique-width este limitat la o constantă. În linii mari, aceste grafuri mai sus menţionate sunt cele care pot fi reduse la arbori, sau la altele clase de bază, folosind descompunerea modulară. Aceste rezultate pot fi regăsite în Capitolul 4 al acestei lucrări.

Tăieturile (splits) ale unui graf sunt un fel de generalizare a modulelor. Există şi unele relaţii între tăieturile unui graf şi parametrul clique-width. Am demonstrat unele relaţii dintre tăieturile unui graf şi lanţurile de creştere, pe baza carora am obţinut algoritmi în timp cvasi-liniar pentru calcularea unui cuplaj maxim în şi mai multe subclase de grafuri ale căror parametru cliquewidth este limitat la o constantă, din care fac parte grafurile ereditare după distanţă. Aceste rezultate pot fi regăsite în Capitolul 5 al acestei lucrări.

Merită menţionată existenţa unor algoritmi destul de simpli pentru calcularea în timp liniar a descompunerii modulare şi a descompunerii unui graf prin tăieturile. Aşadar, rezultatele prezentate în Capitolele 4 şi 5 pot fi aplicate şi pentru prelucrarea reţelelor reale.

Din păcate, nu s-au putut găsi relaţii la fel de interesante, din punct de vedere algoritmic, între lanţurile de creştere si alte structuri combinatorice care sunt legate şi cu parametrul clique-width. Din acest motiv, pentru cazul general, s-a considerat o abordare total diferită, bazată pe logica grafurilor. Întro serie de articole întemeiatoare (cu unii co-autori), pe baza căreia dânsul a obţinut recent premiul Nerode, Courcelle a demonstrat printr-altele rezultate că orice problemă de optimizare care poate fi exprimată întro extensie a logicii monadice de ordinul doi (pe scurt $CMSO_1$, a se vedea Capitolul 2 pentru o definiţie precizǎ) poate fi rezolvată în timp liniar pentru grafurile ale căror parametru cliquewidth este limitat la o constantă[7]. Problema cuplajului maxim nu poate fi exprimată cu această logică! Totodată, o problemă strâns legată: ce este existenţa unui certificat al lui Tutte, poate fi ea exprimată în $CMSO_1$. Am prezentat un algoritm divide et impera pentru calcularea unui cuplaj maxim întrun graf, bazat pe calcularea unor certificate ale lui Tutte în mai multe subgrafuri. Dându-se un graf cu n vârfuri şi m muchii, al cărui

[7] Teorema de optimizare a lui Courcelle presupune pre-calcularea unei k-expresie clique-width (clique-width k-expression), ceea ce este o reprezentare arborescentă doveditoare că valoarea parametrului clique-width este cel mult egală cu k. Acest aspect tehnic este detaliat mai mult în capitolele tehnice ale acestei lucrării.

parametru clique-width este cel mult egal cu k, timpul de rulare al acestui algoritm este în Õ(f(k)·(n+m)), pentru o anumită funcţie f neelementară. Aceste rezultate pot fi regăsite în Capitolul 6 al acestei lucrări.

Existenţa unui algoritm în timp cvasi-liniar care depinde doar *polinomial* de parametrul clique-width rămâne o problemă deschisă.

B.4.2 Contribuţii pentru probleme cu distanţe (Partea II)

Reamintim că *diametrul* unui graf conex este egal cu distanţa maximală dintre două vârfuri oarecare. – Pentru grafurile neconexe, putem defini diametrul ca fiind distanţa maximă dintre două vârfuri în aceaşi componentă conexă. – De asemenea, reamintim că *indicele Wiener* este egal cu suma distanţelor întrun graf, ceea ce este, după normalizare, distanţa medie întrun graf. Numele acestui indice topologic este cel al chimistului Harry Wiener, care a evidenţiat o corelaţie dintre acesta şi punctul de fierbere al unor alcani.

Un centru întrun graf este orice vârf care minimizează distanţa lui maximă până la celelalte vârfuri. Distanţa maximă dintre centru şi un alt vârf se numeşte şi raza grafului. De asemenea, o mediană întrun graf este orice vârf care minimizează distanţa lui medie până la celelalte vârfuri. – Această terminologie a fost introdusă anterior în Secţiunea 1.3. – Centrele şi medianele sunt exact vârfurile ale căror indice de centralitate, după excentricitate şi după apropiere (*closeness*) sunt maximizate – aceşti indici fac parte din metodele de clasificare cele mai populare ale nodurilor întro reţea[8].

Din motivele mai sus expuse, complexitatea problemelor calculării diametrului, razei, indicelui Wiener si a centrelor şi medianelor întrun graf au fost studiate foarte mult în literatură. Acestea pot fi toate rezolvate în timp O(nm) pentru grafurile cu n vârfuri şi m muchii, prin calcularea matricei distanţelor. Totodată, sub ipoteza timpului exponenţial puternic, acest timp de rulare este cvasi-optim. A se vedea şi Capitolul 7.

Existenţa unor algoritmi în timp (cvasi) liniar pentru aceste probleme restricţionate la unele clase de grafuri a fost foarte mult studiată şi în literatură. Totodată, rezultatele obţinute până acum au fost destul de limitate. Spre exemplu, cu ocazia conferinţei organizate cu ocazia implinirii sale a 60 de ani, Chepoi a observat că sunt puţine clase de grafuri pentru care ştim să calculăm mediana în timp liniar, ceea ce include algoritmul dânsului cu co-autori pentru grafurile median, şi algoritmul meu recent pentru grafurile dual cordal.

---

[8] O astfel de metodă alternativă si mult mai populară este cea bazată pe indicii de centralitate după mijlocire (*betweenness*). Din păcate, studiul al complexităţii acestui indice este mai puţin avansat.

În cele ce urmează, se discută doar despre diametrul, pentru care există o literatură mai amplă decât cea pentru celelalte probleme considerate. Una dintre primele rezultate de anvergură pentru calcularea diametrului a fost algoritmul lui Olariu pentru grafurile interval. Ceea ce este interesant este că prin rezultatele prezentate în această teză de abilitare, se evidentiază faptul că sunt multe proprietăţi ale grafurilor interval care pot fi exploatate pentru o calculare eficientă a diametrului în unele clase de grafuri mai mari. Aceşti algoritmi rulează în timp sub-cvadratic, dar nu neapărat în timp liniar. Subiectul existenţei unor algoritmi în timp sub-cvadratic pentru calcularea diametrului a atras atenţia algoritmicienilor doar în ultimii ani, lucrările mai vechi fiind excluziv despre dezvoltarea unor algoritmi în timp liniar sau cvasi-liniar. Această schimbare de paradigmă se datorează marginii inferioare pentru calcularea diametrului, sub ipoteza timpului exponenţial puternic, ceea ce a fost demonstrată recent. Un prim rezultat de importanţă a fost algoritmul lui Cabello pentru calcularea diametrului unui graf planar cu n vârfuri în timp $\tilde{O}(n^{11/6})$. Abordarea lui a fost aplicată si la mai multe probleme de grafuri care sunt legate cu distanţe. În contrast, sunt unele clase de grafuri foarte simple, din punct de vedere structural şi metric, pentru care calcularea diametrului nu se poate face decât în timp cvadratic (sub unele ipoteze de complexitate standarde). Acestea includ grafurile bipartite şi grafurile split. Deocamdată, nu există o caracterizare satisfăcătoare despre clasele de grafuri ale căror problema diametrului poate fi rezolvată în timp sub-cvadratic.

Rezultate. Întro serie de articole recente (alături de mai mulţi co-autori), am contribuit la înţelegerea mult mai bună despre complexitatea calculării diametrului în unele clase de grafuri foarte generale. Rezultatele noastre pot fi generalizate parţial la mai multe probleme legate cu distanţe. Ceea ce este remarcabil la acestea este nivelul lor de generalitate. Spre exemplu, printro proprietate comună a grafurilor interval şi a grafurilor planare, se poate deduce existenţa unui algoritm eficient (în timp sub-cvadratic) pentru calcularea diametrului în ambele clase de grafuri.

Mingea centrată la vârful v cu rază k este mulţimea vârfurilor care stau la distanţa cel mult k raportat la v. Putem defini *hipergraful mingiilor* unui graf G ca fiind mulţimea mingiilor acestui graf (cu centrul şi raza oarecare). Studiul proprietăţilor acestui hipergraf are sens în cadrul studiului complexităţii unor probleme despre grafurile legate cu distanţe. Hipergrafurile au fost considerate ca fiind un model combinator pentru problemele de cerinţe de interval în spaţiile metrice. Din acest motiv, proprietăţile hipergrafurilor au fost studiate foarte mult în domeniul Geometriei Computaţionale. Un merit al aceastei lucrări este de a transfera o parte din aceste rezultate la problemele pentru grafuri. Abordarea propusă se bazează şi pe Teoria Metrică Grafurilor, în care se studiază clasele de grafuri ale căror funcţie a distanţelor are unele proprietăţi asemănătoare cu spaţiile metrice clasice, precum spaţiile Euclidiene şi spaţiile Hiperbolice.

Grafurile Helly sunt exact cele ale căror hipergraf al mingiilor îndeplineşte proprietatea lui Helly: dându-se o mulţime de mingi în aşa fel încât fiecare două mingi se intersectează, intersecţia tuturor mingiilor mulţimii este nevidă. Clasa grafurilor Helly este una dintre cele mai studiate în Teoria Metrică Grafurilor. Întradevăr, orice graf este subgraful izometric al unui graf Helly. Mai mult, reţelele complexe, deşi în general nu sunt grafuri Helly, îndeplinesc o variantă puţin mai slabă a proprietăţii lui Helly. Sunt şi unele clase de grafuri cu unele aplicaţii practice, precum grafurile interval, care sunt grafuri Helly. Am iniţiat studiul complexităţii problemelor de calculare a diametrului, razei şi medianelor întrun graf Helly. Acest studiu a fost realizat în colaborare cu F. Dragan şi H. Guarnera. În mai multe cazuri, am obţinut unii algoritmi eficienţi folosind o strategie simplă de căutare locală. Aceste rezultate pot fi regăsite în Capitolul 8 al acestei lucrări.

Dimensiunea Vapnik-Chervonenkis (pe scurt, VC-dimension) a fost studiată în foarte multe domenii, precum învăţarea automată. De asemenea, a fost studiată în Teoria Grafurilor. S-a definit parametrul *distance VC-dimension* al unui graf G ca fiind dimensiunea Vapnik-Chervonenkis asociată cu hipergraful mingiilor lui G. Merită menţionat că parametrul distance VC-dimension este limitat la o constantă în următoare clasele de grafuri: grafurile interval, grafurile ale căror parametru clique-width (sau tree-width) este limitat la o constantă, grafurile planare şi, mai general, toate grafurile care pot fi încorporate întro suprafaţă al carui gen este limitat la o constantă, în aşa fel încât muchiile se pot intersecta doar în varfuri. Am prezentat unii algoritmi în timp sub-cvadratic pentru calcularea diametrului în toate clasele de grafuri mai sus menţionate. În special, am generalizat algoritmul lui Cabello pentru grafurile planare la toate clasele de grafuri care exclud un minor fixat. Aceste rezultate au la bază o proprietate a hipergrafurilor cu dimensiunea Vapnik-Chernonenkis limitată la o constantă demonstrată anterior. Acestea pot fi regasite în Capitolul 9 al acestei lucrări.

În sfârşit, se consideră existenţa unor algoritmi în timp cvasi-liniar pentru calcularea diametrului, o provocare mult mai mare. Recent, au fost prezentaţi unii algoritmi parametrizaţi pentru calcularea diametrului şi al indicelui Wiener al unui graf, în cazul special în care parametrul tree-width este limitat la o constantă. S-a demonstrat şi că dependinţa exponenţială în tree-width a timpului de rulare este optimă, sub ipoteza timpului exponenţial puternic. Aceste rezultate sunt bazate pe o structură de date introdusă anterior de către Bentley pentru cerinţele de interval în bazele de date. S-a observat că această structură de date poate fi aplicată şi la rezolvarea unor probleme ale grafurilor care sunt legate cu distanţe. În linii mari, graful este deconectat recursiv cu unele mulţimi separatoare cu k vârfuri, şi la fiecare pas, vârfurile sunt reduse la un k-tuplu de distanţe. Am propus unele metode asemănătoare, dar folosind alte structuri decât mulţimile separatoare, precum: tăieturile (cu aplicaţii la calcularea eficientă a diametrului în grafurile ale căror parametru clique-width este limitat la o constantă), reprezentările unui graf cu intervale, şi hub labeling. Aceste rezultate pot fi regăsite în Capitolul 10 al acestei lucrări.

O perspectivă de cercetare interesantă ar fi unificarea rezultatelor în Capitolele 8 şi 9, şi a mai multor rezultate din literatură, printrun studiu al proprietăţilor asemănătoare cu cea a lui Helly.

B.5 Organizarea capitolelor

În Capitolul 2, pot fi regăsite unele rezultate preliminare despre Teoria Complexităţii şi Algoritmica Grafurilor. Câteva din aceste rezultate sunt contribuţii individuale, sau produsul unei colaborări cu mai mulţi co-autori. În capitolele următoare, ne vom referi mereu la aceste rezultate.

Partea I a tezei de abilitare este dedicată la un studiu parametrizat despre problema calculării unui cuplaj maxim întrun graf. Se consideră în special parametrul clique-width, şi alţi parametrii legaţi cu acesta. Capitolul 3 este un rezumat despre teoria cuplajelor întrun graf, în care se menţionează şi noţiunile mai generale de b-cuplaj şi de f-factor. Sunt introduse şi conceptele cheie precum: *lanţuri de creştere* şi *Matricea lui Tutte*. În Capitolele 4 şi 5, se prezintă unii algoritmi pentru calcularea unui cuplaj maxim în timp cvasi-liniar, pentru mai multe subclase de grafuri în care parametrul clique-width este limitat la o constantă. Aceste rezultate algoritmice au la bază studiul relaţiilor dintre lanţurile de creştere şi unele structuri combinatorice precum modulele şi tăieturile. În Capitolul 6, se prezintă o abordare mai generală, bazată pe logica grafurilor, care poate fi aplicată la orice graf cu parametrul clique-width limitat la o constantă.

În Partea II a tezei de abilitare, se prezintă mai mulţi algoritmi în timp subcvadratic pentru calcularea exactă a diametrului, şi respectiv a altor probleme legate cu distanţe întrun graf, pentru mai multe clase de grafuri generale. Rezultatele despre complexitatea acestor probleme, pentru grafurile oarecare, sunt reamintite în Capitolul 7. Apoi, în următoarele trei capitole, se prezintă mai mulţi algoritmi generici pentru aceste probleme, în mai multe clase de grafuri, care rulează în timp sub-cvadratic. Aceste rezultate au la bază unele metode inspirate din domeniul Geometriei Computaţionale. În special, un studiu parametrizat despre aceste probleme este prezentat în Capitolele 8 şi 9, parametrii consideraţi fiind numărul Helly şi dimensiunea Vapnik-Chervonenkis a unui graf. Acest studiu este, fără îndoială, rezultatul principal din cadrul aceastei abilitări. În Capitolul 10 se consideră şi o tehnică pentru a rezolva unele cerinţe de interval în spaţiile Euclidiene, a cărei utilitatea în Algoritmica Grafurilor a fost dovedită pentru prima dată de către Cabello şi Knauer. Se prezintă mai multe aplicaţii noi ale aceastei tehnici.

Concluzia lucrării se regaseşte în Capitolul 11. Se menţionează şi nişte perspective despre rezultatele obţinute, şi direcţii de cercetare.