



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI
MINISTERUL MUNCII, FAMILIEI,
PROTECȚIEI SOCIALE ȘI
PERSOANELOR VÂRSTNICE
AMPOSDRU



Fondul Social European
POSDRU 2007-2013



Instrumente Structurale
2007-2013



MINISTERUL
EDUCAȚIEI
NAȚIONALE

OIPOSDRU



Universitatea
POLITEHNICA
din București

FONDUL SOCIAL EUROPEAN

Investește în oameni!

Programul Operațional Sectorial pentru Dezvoltarea Resurselor Umane 2007 – 2013

Proiect POSDRU/107/1.5/S/76903 – *Formarea viitorilor cercetători-experti prin programe de burse doctorale (EXPERT)*



UNIVERSITATEA **POLITEHNICA** DIN BUCUREȘTI

Facultatea de Automatică și Calculatoare

Departamentul de Calculatoare

Nr. Decizie Senat 228 din 13.03.2014

TEZĂ DE DOCTORAT

REZUMAT

Detectarea Propagării Opiniilor în Grafuri cu Conținut Dinamic

Detection of Opinion Propagation in Graphs with Dynamic Content

Autor: As.ing. Dumitru-Clementin Cercel

Conducător de doctorat: Prof.dr.ing. Ștefan Trăușan-Matu

COMISIA DE DOCTORAT

Președinte	Prof.dr.ing. Adina-Magda Florea	de la	Universitatea Politehnica din București
Conducător de doctorat	Prof.dr.ing. Ștefan Trăușan-Matu	de la	Universitatea Politehnica din București
Referent	Prof.dr. Constanța-Nicoleta Bodea	de la	Academia de Studii Economice din București
Referent	Prof.dr.ing. Florin Rădulescu	de la	Universitatea Politehnica din București
Referent	Cercetător Științific I Dr. Costin Pribeanu	de la	Institutul Național de Cercetare-Dezvoltare în Informatică – ICI București

București, 2014

CUPRINS

CAPITOLUL 1. Introducere	4
1.1. Motivație	4
1.2. Obiectivele Tezei.....	4
1.3. Structura Tezei.....	5
 CAPITOLUL 2. Mineritul Opiniilor	8
2.1. Opinii	8
2.2. Problema Mineritului Opiniilor	8
2.3. Mineritul Opiniilor bazat pe Aspecte	9
2.4. Tipuri de Analiză a Sentimentelor	9
2.5. Vocabulare de Opinii	9
 CAPITOLUL 3. Grafuri cu Conținut Dinamic	11
3.1. Concepte de Bază	11
3.2. Rețele Modelate sub Formă de Grafuri.....	11
3.3. Rețele de Socializare Online Modelate sub Formă de Grafuri	11
3.4. Modele de Grafuri pentru Rețelele de Socializare Online	12
 CAPITOLUL 4. Propagarea Opiniilor în Rețelele de Socializare Online	13
4.1. Modelarea Propagării Opiniilor în Rețelele de Socializare Online	13
4.2. Discuții	14
 CAPITOLUL 5. Provocări în Cercetarea Mineritului Opiniilor dintr-o	15
Perspectivă a Prelucrării Limbajului Natural	
5.1. Preprocesarea de Texte	15
5.2. Extragerea Cuvintelor de Opinie	15
5.3. Detectarea Polarității Cuvintelor de Opinie	16
5.4. Rolul Etichetării cu Părți de Vorbire a Cuvintelor în Mineritul Opiniilor	16
 CAPITOLUL 6. Analiza Propagării Opiniilor la Nivel de Utilizator în Firele	17
de Discuție din Forumuri	
6.1. Introducere	17
6.2. Analiza Firelor de Discuție din Forumuri	17
6.3. Formularea Problemei	18
6.4. Arhitectura Abordării Propuse.....	20
6.5. Experimente	21

CAPITOLUL 7. Analiza Propagării Opiniilor la Nivel de Postare în Firele de	25
Discuție din Forumuri	
7.1. Definirea Problemei	25
7.2. Arhitectura Abordării Propuse	25
7.2.1. Eliminarea Postărilor care Conțin numai Fapte	26
7.2.2. Identificarea Sentimentului la Nivel de Postare	27
7.2.3. Eliminarea Postărilor cu Sentiment Neutru	28
7.2.4. Agregarea de Noduri Părinte-Copil	29
7.2.5. Agregarea de Noduri Frați	30
 CAPITOLUL 8. Analiza Influenței Opiniilor în Firele de Discuție din	33
Forumuri	
8.1. Introducere	33
8.2. Formularea Problemei	33
8.3. Arhitectura Abordării Propuse	34
8.4. Experimente	34
 CAPITOLUL 9. Concluzii și Direcții Ulterioare	38
9.1. Contribuții Personale	39
9.2. Direcții Ulterioare de Cercetare	40
 LISTA DE PUBLICAȚII	42
 REFERINȚE	43

CAPITOLUL 1

Introducere

1.2. Motivație

Societatea contemporană se confruntă cu fenomenul de interacțiune între oameni în mediile sociale, fapt ce are implicații majore pentru propagarea opiniilor în astfel de medii. În prezent, procesul propagării opiniilor se deosebește fundamental de cel derulat în perioadele anterioare: prin rapiditate și o largă răspândire. Astfel, s-a ajuns ca, în epoca Internetului, opiniile utilizatorilor să fie influențate de comentariile altor oameni, chiar și atunci când nu se cunosc personal.

Procese de propagare în diverse medii sociale (bloguri, rețele de socializare online, forumuri de discuții) sunt variate: propagarea de informații, opinii, zvonuri, inovații tehnologice, știri etc. Toate aceste procese prezintă o serie de caracteristici, ceea ce impune dezvoltarea unor metode specifice pentru a le studia. Mai mult, mediile sociale pot fi modelate utilizând diferite tipuri de grafuri, iar apoi analizate aplicând teoria grafurilor. Aceste tipuri de grafuri se construiesc pe baza relațiilor stabilite între utilizatori sau a acțiunilor utilizatorilor în mediile sociale rezultate ca urmare a interacțiunilor dintre ei.

Analiza propagării opiniilor în mediile sociale este un subiect de cercetare relativ nou, determinat atât de creșterea popularității mediilor sociale, cât și de dezvoltarea domeniului mineritului opiniilor. Acest subiect este unul de mare actualitate și importanță științifică deoarece fenomenul propagării opiniilor în mediile sociale are mai multe aplicații posibile în lumea reală, de exemplu în campaniile electorale și în marketing.

1.2. Obiectivele Tezei

Obiectivul principal al acestei teze este de a cerceta problema propagării opiniilor în mediile sociale. Concret, noi vom studia problema propagării opiniilor în două tipuri de medii sociale: rețelele de socializare online și firele de discuție din forumuri. După cunoștințele noastre, problema propagării opiniilor în firele de discuție din forumuri nu a fost abordată în cercetarea anterioară. În plus, cercetarea cu privire la propagarea opiniilor în rețelele de socializare online nu a ținut seama de mineritul opiniilor. Prin urmare, în metodele pe care le propunem în această teză, vom combina domeniile științifice: mineritul opiniilor cu analiza propagării opiniilor în firele de discuție din forumuri.

De asemenea, există mai multe obiective secundare ale tezei. În continuare, vom formula aceste obiective secundare sub formă de întrebări de cercetare:

- *Cum putem modela propagarea opiniilor în rețelele de socializare online?*
- *Cum se propagă opiniile de-a lungul timpului în firele de discuție din forumuri?*
- *Cum se influențează de-a lungul timpului opiniile exprimate de participanți în firele de discuție din forumuri?*

- Cum pot fi extrase opiniile din mesajele trimise de utilizatori în mediile sociale?
- Cum sunt influențate rezultatele procesului de etichetare gramaticală cu partea de vorbire corespunzătoare a fiecărui cuvânt dintr-un text, atunci când se folosește o metodă de învățare supervizată, iar dimensiunea corpusului de antrenare variază?
- Cum sunt influențate rezultatele procesului de etichetare gramaticală cu partea de vorbire corespunzătoare a fiecărui cuvânt dintr-un text, atunci când se folosește o metodă de învățare supervizată și ambele corpusuri (de antrenare, respectiv de testare) nu sunt din același domeniu?

Obiectivele acestei teze au fost rezolvate într-unul sau mai multe capitole, așa cum se arată în Tabelul 1.1.

Tabel 1.1: Structura Tezei.

Obiectivele Tezei	Capitolele Tezei
Mineritul opiniilor din texte	Capitolul 2, Capitolul 5
Problema propagării opiniilor în rețelele de socializare online	Capitolele 3 – 4
Problema propagării opiniilor în firele de discuție din forumuri	Capitolele 6 – 8

1.3. Structura Tezei

După capitolul introductiv, conținutul acestei teze este structurat astfel:

Capitolul 2 analizează problema mineritului opiniilor din texte. Pentru a oferi o bază de plecare în cercetarea mineritului opiniilor, mai întâi trebuie definite și corelate între ele noțiunile specifice acestui domeniu. Astfel, vom analiza conceptul de *opinie* din mai multe perspective: limbajul colocvial, limbajul tehnic și din punct de vedere computațional. După clarificarea acestui concept, vom da o definiție formală problemei mineritului opiniilor. De asemenea, vom evidenția diferențele dintre problema mineritului opiniilor în general și problema mineritului opiniilor bazat pe aspecte.

O altă secțiune prezintă tipurile de granularitate studiate în analiza sentimentelor: la nivel de document, la nivel de propoziție și la nivel de cuvânt. La sfârșitul acestui capitol, vom descrie patru vocabulare de opinii: SentiWordNet [1], Micro-WNOp [2], vocabularul de subiectivitate MPQA [3] și vocabularul de opinii creat de Bing Liu [4].

Pentru a înțelege specificul datelor utilizate în cercetarea noastră, adică mediile sociale cu un conținut dinamic, explicații necesare sunt prevăzute în **Capitolul 3**. Vom începe cu o prezentare a conceptelor de bază utilizate în teoria grafurilor. Vom continua cu o descriere de ansamblu a rețelelor care sunt modelate sub formă de grafuri.

Secțiunea 3.3 cuprinde o descriere a rețelelor de socializare online, subliniind caracteristicile lor. În continuare, vom descrie mai multe modele de grafuri pentru rețelele de socializare online. Aceste modele au fost folosite pentru a simula propagarea opiniilor în rețelele de socializare online, așa cum este arătat în Capitolul 4.

Capitolul 4 este dedicat problemei propagării opiniilor în rețelele de socializare online. Mai multe modele au fost propuse în literatura de specialitate. Am clasificat modelele care studiază problema propagării opiniilor în două categorii: modele bazate pe grafuri nedirecționate pentru detectarea propagării opiniilor și modele bazate pe grafuri direcționate pentru detectarea propagării opiniilor.

În continuare, vom analiza pentru cele două categorii, principalele modele de propagare de opinii elaborate în literatura științifică: modelul Voter [5], modelul Sznajd [6], modelul Deffuant [7], modelul Hegselmann-Krause [8], modelul Independent Cascade [9] și modelul Linear Threshold [9]. În cele din urmă, vom discuta despre dezavantajele identificate în aceste modele.

Capitolul 5 conține o analiză a dificultăților legate de problema mineritului opiniilor din texte, dintr-o perspectivă a prelucrării limbajului natural (PLN). În acest capitol vom prezenta sarcinile specifice prelucrării limbajului natural, care influențează precizia rezultatelor mineritului opiniilor. Deși îndelung studiate în comunitatea de cercetare PLN, aceste sarcini nu sunt complet rezolvate. Concret, noi vom împărți problema mineritului opiniilor în trei etape: etapa de preprocesare de texte în Secțiunea 5.1, etapa de extragere a cuvintelor de opinie în Secțiunea 5.2 și etapa de detectare a polarității cuvintelor de opinie în Secțiunea 5.3.

La sfârșitul acestui capitol, vom prezenta un studiu de caz cu privire la rolul etichetării cu părți de vorbire a cuvintelor în mineritul opiniilor. Acest studiu de caz răspunde la două dintre întrebările secundare de cercetare formulate în secțiunea 1.2.

În **Capitolul 6** prezentăm o metodă proprie pentru a studia problema propagării opiniilor în firele de discuție din forumuri. Analiza propagării se face la nivel de utilizator.

Acest capitol este structurat după cum urmează. Capitolul începe cu o prezentare generală a metodei propuse, urmată de o descriere formală a firelor de discuție. În secțiunea 6.3, am definit în mod formal problema propagării opiniilor la nivel de utilizator în firele de discuție din forumuri.

În secțiunea 6.4, am detaliat sistemul propus de noi pentru a rezolva această problemă. Astfel, trei etape sunt descrise în următoarea ordine: extragerea cuvintelor de opinie din postările participanților, construirea vocabularului de termeni substantivali, precum și identificarea sentimentelor exprimate de cuvintele de opinie. Sentimentele cuvintelor de opinie sunt determinate cu ajutorul a patru vocabulare de opinii. Rezultatele experimentale sunt prezentate în secțiunea 6.5.

În **Capitolul 7** prezentăm o altă metodă proprie pentru a studia problema propagării opiniilor în firele de discuție din forumuri. În acest caz, analiza propagării se face la nivel de postare. Această metodă permite vizualizarea, într-un mod simplificat, a grafului asociat firului de discuție dintr-un forum. În acest scop, am creat patru algoritmi. Mai mult decât atât, am propus o formulă pentru a identifica sentimentul exprimat la nivel de postare.

Prima secțiune a acestui capitol cuprinde o introducere în metoda propusă de noi pentru a studia problema propagării opiniilor în firele de discuție din forumuri. Vom continua, în secțiunea 7.2, cu descrierea arhitecturii sistemului care implementează metoda propusă. Detaliem implementarea fiecărei etape a acestei arhitecturi în cadrul unei subsecțiuni distincte: etapa de eliminare a postărilor care conțin numai fapte, etapa de identificare a sentimentului la nivel de postare, etapa de eliminare a postărilor cu sentiment neutru, etapa de agregare de noduri părinte-copil și etapa de agregare de noduri frați.

În **Capitolul 8** prezentăm o metodă proprie pentru a studia influența dintre participanții care și-au exprimat opiniile într-un fir de discuție dintr-un forum. Un participant pe un fir de discuție din forum este influențat de alți participanți la discuție dacă acesta trimite, de-a lungul timpului, postări de polaritate opusă primei postări trimise de el. În cazul afirmativ, vom determina postarea / postările care l-au influențat cel mai mult pe un participant să-și schimbe opinia / opiniile. Pentru a calcula similaritatea semantică între două postări, vom folosi analiza semantică latentă [10, p. 243].

Acest capitol este structurat după cum urmează. În secțiunea 8.1, este făcută o introducere în metoda propusă de noi. Secțiunea 8.2 oferă un cadru teoretic pentru analiza influenței opiniilor în firele de discuție din forumuri. În secțiunea 8.3, vom prezenta arhitectura sistemului care implementează metoda. Rezultatele experimentale sunt prezentate în secțiunea 8.4.

Teza se încheie cu **Capitolul 9**, care conține concluziile, urmate de o trecere în revistă a principalelor noastre contribuții științifice. În cele din urmă, vom prezenta câteva direcții pentru activitatea de cercetare viitoare.

CAPITOLUL 2

Mineritul Opiniilor

2.1. Opinii

Dintr-o perspectivă computațională, Kim și Hovy [11] definesc conceptul de *opinie* cu ajutorul a patru cuvinte-cheie (titular opinie, punct de vedere, topică, sentiment): „Titularul opiniei este cel care are un punct de vedere cu privire la o anumită topică, în multe cazuri putându-i asocia un sentiment”. *Topica* (de asemenea, numită *obiect*, *entitate* sau *țintă* în literatura de specialitate) despre care titularul opiniei își formează un *punct de vedere* poate fi reprezentată de un produs, o persoană sau un eveniment [12]. *Titularul opiniei* este individul sau organizația care și-a exprimat punctul de vedere [12]. *Sentimentele* exprimate de către opinii pot fi: pozitive, negative sau neutre.

Pentru a înțelege această definiție, vom folosi următorul exemplu: „Eu cred că aparatul de fotografiat este frumos”. În această propoziție, titularul opiniei este „Eu”, entitatea țintă este „aparatul de fotografiat”, iar sentimentul opiniei este pozitiv. Însă, definiția lui Kim și Hovy nu ține seama de faptul că nu întotdeauna o persoană își exprimă opinia cu privire la entitatea în sine, ci sunt situații când opiniile sunt exprimate în legătură cu aspectele entității.

O altă abordare a conceptului de *opinie* este adoptată de către Liu [10, p. 463] care a definit riguros un model ce ia în considerare aspectele entității: „o opinie este un cvintuplu de elemente ($e_i, a_{ij}, o_{ijkl}, h_k, t_l$), unde e_i este numele unei entități, a_{ij} este un aspect al entității e_i , o_{ijkl} este orientarea opiniei cu privire la aspectul a_{ij} al entității e_i , h_k este titularul opiniei, iar t_l este momentul de timp în care opinia este exprimată de h_k ”.

2.2. Problema Mineritului Opiniilor

Mineritul opiniilor este un domeniu de cercetare de mare actualitate. Dezvoltarea acestui domeniu se datorează popularității în continuă creștere a Web-ului, care, de altfel, generează o cantitate imensă de date electronice [13, p. 7]. Mineritul opiniilor este un domeniu de cercetare în care sunt aplicate sarcini specifice prelucrării limbajului natural (PLN). Principala deosebire este că PLN cuprinde sarcini ce abordează întreaga semantică a unui text, în timp ce mineritul opiniilor se limitează la unele aspecte semantice ale textului, precum entitatea țintă, opiniile cu privire la entitatea țintă, sentimentele acestor opinii [13, p. 13].

Fiind date un corpus de texte și o anumită entitate țintă, problema mineritului opiniilor presupune două etape: în primul rând, identificarea din texte a opiniilor cu privire la entitatea țintă (etapa extragerii de opinii), iar în al doilea rând, determinarea orientării semantice a fiecărei opinii, care poate fi pozitivă, negativă sau neutră (etapa detectării polarității opiniilor). În plus, pentru a rezolva aceste două etape, adeseori este necesară o etapă de preprocesare a datelor de intrare. Etapa de preprocesare este necesară deoarece datele de intrare trebuie pregătite în mod corespunzător pentru prelucrările ulterioare. Deși pot fi folosite metode specifice pentru a rezolva fiecare dintre cele trei etape, abordările propuse pentru problema mineritului opiniilor trebuie să țină cont de faptul că există unele sarcini PLN care influențează rezultatele.

Într-o altă concepție [14], mineritul opiniilor implică mai multe etape. Procesul mineritului opiniilor începe cu o etapă de clasificare binară pentru a distinge între propozițiile obiective și cele subiective. Numai propozițiile subiective sunt luate în considerare mai departe. Astfel, pentru fiecare propoziție subiectivă, tipul de polaritate este determinat. Acest algoritm nu extrage toate opiniile exprimate în texte deoarece subiectivitatea nu este echivalentă cu exprimarea de opinii [13, p. 27]. Prin urmare, mineritul opiniilor și analiza subiectivității sunt domenii parțial similare.

2.3. Mineritul Opiniilor bazat pe Aspecte

O entitate (adică un produs, o organizație sau un eveniment) poate fi caracterizată de o mulțime de componente și de o mulțime de attribute, astfel încât fiecare componentă poate fi împărțită, la rândul său, în subcomponentele și attributele ei și așa mai departe. Toate aceste diviziuni sunt denumite *aspecte* sau *trăsături* [12]. De exemplu, o „cameră digitală” este o entitate ce are o mulțime de componente („obiectiv”, „baterie” etc.), precum și o mulțime de attribute („calitatea imaginii”, „dimensiune” etc.). Componenta „baterie” are, de asemenea, o mulțime de attribute („durată”, „dimensiune” etc.).

Putem extinde problema mineritului opiniilor definită în general la problema mineritului opiniilor bazat pe aspecte, care, potrivit lui Bing Liu [10, p. 480-481], constă în determinarea tuturor cvintuplurilor ($e_i, a_{ij}, o_{ijkl}, h_k, t_l$) existente într-un anumit corpus de texte. Concret, această problemă presupune două etape: în primul rând, extragerea componentelor, subcomponentelor și atributelor entității despre care opiniile sunt exprimate, iar în al doilea rând, determinarea sentimentului fiecărei opinii.

2.4. Tipuri de Analiză a Sentimentelor

Fiind dat un document în care o persoană evaluează o entitate, exprimându-și opiniile cu privire la aceasta, problema clasificării sentimentului la nivel de document urmărește să determine sentimentul de ansamblu al opiniilor exprimate de acea persoană [10, p. 30]. Tehnicile care rezolvă acest tip de problemă sunt în general bazate pe învățarea supervizată [15], deși există, de asemenea, unele metode nesupervizate [16].

La nivel de propoziție, problema clasificării sentimentului își propune să determine dacă o propoziție subiectivă exprimă un sentiment pozitiv, negativ sau neutru. Această problemă presupune două sarcini [10, p. 475]: rezolvarea sarcinii de clasificare a subiectivității, urmată de rezolvarea sarcinii de clasificare a propozițiilor subiective.

Metodele de identificare a sentimentului exprimat de cuvintele de opinie pot fi manuale, bazate pe corpusuri sau bazate pe dicționare [13, p. 90]. Abordarea manuală este dificilă deoarece necesită mult timp pentru rezolvare. Cu toate acestea, abordarea manuală poate fi aplicată înainte sau după celelalte două metode pentru a îmbunătăți precizia rezultatelor obținute [4].

2.5. Vocabulare de Opinii

Vocabularele de opinii sunt resurse importante ce permit identificarea polarității cuvintelor de opinie. Câteva vocabulare de opinii sunt disponibile cu ajutorul cărora putem clasifica cuvintele de opinie într-una din următoarele categorii de sentimente: pozitive, negative sau neutre. Dintre vocabularele de opinii existente, noi am selectat patru: SentiWordNet [1], Micro-WNOp [2],

vocabularul de subiectivitate MPQA [3], precum și vocabularul de opinii creat de Bing Liu [4]. În Tabelul 2.1, vom enumera principalele diferențe între acestea.

Tabelul 2.1: Vocabulare de Opinii.

Vocabular de Opinii	Dimensiunea Vocabularului de Opinii	Categorii de Sentimente	Trăsăturile Vocabularului de Opinii
SentiWordNet 3.0 (bazat pe WordNet ¹ 3.0)	155.287 de cuvinte din 117.659 de seturi de sinonime	Fiecare cuvânt are un scor cuprins între 0 și 1 pentru fiecare categorie de sentimente: pozitive, negative sau neutre.	• Este generat în mod semiautomat; • Distinge între părțile de vorbire ale cuvintelor de opinie.
Micro-WNOp (bazat pe WordNet 2.0)	1.960 de cuvinte din 1.105 de seturi de sinonime	Fiecare cuvânt are un scor cuprins între 0 și 1 pentru fiecare categorie de sentimente: pozitive, negative sau neutre.	• Este generat în mod manual; • Distinge între părțile de vorbire ale cuvintelor de opinie.
Vocabularul de Subiectivitate MPQA	8.221 de cuvinte	Fiecare cuvânt este inclus într-una dintre următoarele categorii de sentimente: pozitive, negative, ambele (pozitive și negative) sau neutre.	• Este generat în mod semiautomat; • Distinge între părțile de vorbire ale cuvintelor de opinie.
Vocabularul de Opinii creat de Bing Liu (bazat pe WordNet 2.0)	6.786 de cuvinte	Fiecare cuvânt este inclus într-una dintre următoarele categorii de sentimente: pozitive sau negative.	• Este generat în mod semiautomat; • Nu distinge între părțile de vorbire ale cuvintelor de opinie.

Un cuvânt de opinie care are mai multe sensuri, poate exprima sentimente diferite. Vocabularul SentiWordNet permite identificarea sentimentului exprimat de fiecare cuvânt de opinie, luând în considerare sensul său. Celelalte trei vocabulare de opinie (Micro-WNOp, vocabularul de subiectivitate MPQA, vocabularul de opinii creat de Bing Liu) nu fac nicio distincție între sensurile unui cuvânt. Mai mult decât atât, unele vocabulare de opinii (SentiWordNet și Micro-WNOp) definesc un scor de sentiment pentru fiecare cuvânt de opinie.

¹ <http://wordnet.princeton.edu/>

CAPITOLUL 3

Grafuri cu Conținut Dinamic

3.1. Concepte de Bază

Considerăm un graf G ca o pereche de mulțimi (V, E) , unde V este mulțimea nodurilor în graf, iar E este mulțimea muchiilor care conectează nodurile în același graf. Într-un graf direcționat, orice muchie din mulțimea E este orientată, direcția orientării fiind indicată de o săgeată. În ceea ce privește grafurile nedirecționate, toate muchiile din mulțimea E sunt neorientate.

3.2. Rețele Modelate sub Formă de Grafuri

Fenomenele lumii reale pot fi reprezentate sub formă de rețele. Potrivit lui Wang [17], conceptul de *rețea* este definit printr-o mulțime de entități cu un conținut specific și o mulțime de legături între entități. În funcție de domeniul de aplicare, există denumiri specifice pentru entitățile și legăturile unei rețele [18]. Rețelele sunt studiate folosind teoria grafurilor deoarece componentele rețelelor pot fi reprezentate cu ajutorul grafurilor. De altfel, teoria grafurilor furnizează o serie de concepte care permit analiza proprietăților rețelelor [19, p. 15].

3.3. Rețele de Socializare Online Modelate sub Formă de Grafuri

O rețea de socializare este definită printr-o mulțime de persoane și o mulțime de relații și interacțiuni între aceste persoane. Exemple clasice de rețele de socializare sunt rețelele de socializare online, cum ar fi Facebook¹, Twitter² sau LinkedIn³. Alte exemple de rețele de socializare sunt rețelele de comunicare (de exemplu, o rețea de apeluri telefonice, o rețea de mesaje electronice etc.) sau rețelele de colaborare între muzicieni, actorii de filme etc.

Într-o rețea de socializare, relațiile dintre membri pot fi de mai multe tipuri: relații de prietenie, relații de rudenie, relații de colaborare, relații de afaceri etc. Rețelele de socializare online permit membrilor să stabilească legături explicite între ei (de exemplu, relațiile de prietenie pe Facebook sau relațiile „de urmărire” / „urmărit” pe Twitter). Aceste relații reprezintă o condiție prealabilă pentru stabilirea interacțiunilor între membrii unei rețele de socializare online. Interacțiunile dintre membrii unei rețele de socializare online sunt determinate de acțiunile pe care aceștia le pot face pe acea rețea: postarea de comentarii, urmărirea activității altor persoane, folosirea butonului „like” etc. Mai mult decât atât, în unele rețele de socializare online, persoanele primesc și trimit mesaje altor membri din rețea. Aceste mesaje pot fi de mai multe tipuri: mesaje care conțin opinii, mesaje care conțin fapte, mesaje care conțin zvonuri, mesaje despre inovații tehnologice.

Relațiile și interacțiunile stabilite între membrii unei rețele de socializare online ne permit să reprezentăm comunitatea de indivizi printr-un graf social $G_s = (V_s, E_s)$, unde V_s este mulțimea

¹ <https://www.facebook.com/>

² <https://twitter.com/>

³ <https://www.linkedin.com/>

nodurilor rețelei (altfel spus, mulțimea de membri ai rețelei) iar E_s este mulțimea muchiilor care conectează nodurile.

Graful social poate fi direcționat sau nedirecționat. Un graf social direcționat este folosit pentru a modela rețelele de socializare online în care relațiile dintre membri nu sunt simetrice, de exemplu relațiile „urmărit” de pe Twitter. În mod analog, un graf social nedirecționat modelează rețelele de socializare online care conțin relații simetrice între membri, de exemplu relațiile de prietenie de pe Facebook.

3.4. Modele de Grafuri pentru Rețelele de Socializare Online

Rețelele de socializare online au proprietăți de structură similare cu proprietățile rețelilor generate utilizând următoarele modele [20, 21]: modelul de rețea de tip Watts-Strogatz (WS) [22] și modelul de rețea de tip Barabási-Albert (BA) [23]. În ambele modele, legăturile dintre noduri sunt simetrice.

În modelul de rețea WS, există inițial o mulțime de noduri dispuse într-o topologie de tip inel, unde fiecare nod este conectat cu cele mai apropiate patru noduri. Cu aceeași probabilitate p , fiecare muchie poate fi refăcută, ceea ce înseamnă că un capăt al muchiei este mutat la un alt nod ales în mod aleator din rețea, astfel încât două condiții să fie îndeplinite: este posibilă doar o singură muchie între oricare două noduri, iar buclele nu sunt permise. Rețelele de tip small-world generate folosind modelul WS au două proprietăți: lungimea de cale medie are o valoare redusă și coeficientul de clusterizare are o valoare mare, ceea ce înseamnă că toate nodurile au aproximativ același grad [17]. Rețelelor de tip small-world le lipsesc două caracteristici fundamentale ale rețelilor de socializare online [17]: adăugarea de noi noduri în rețea de-a lungul timpului și crearea de conexiuni preferențiale între noduri de-a lungul timpului.

Modelul de rețea BA ia în considerare aceste două caracteristici. Fiind dată o mulțime inițială de noduri, la fiecare moment de timp rețeaua crește prin adăugarea unui nod. Nodurile noi stabilesc legături în mod aleator cu nodurile care existau deja în rețea. În comparație cu rețelele de tip small-world, în rețelele de tip scale-free, conexiunile pentru noile noduri din rețea se realizează cu diferite probabilități și mai ales cu nodurile de grad mai mare. Rețelele de tip scale-free generate prin utilizarea modelului BA au următoarele proprietăți: distribuția de grad pentru noduri urmează o lege putere cu valoarea exponent 3, lungimea de cale medie crește în timp, iar coeficientul de clusterizare scade în timp. În figura 3.1 sunt ilustrate trei tipuri de rețele generate folosind modelele WS și BA.

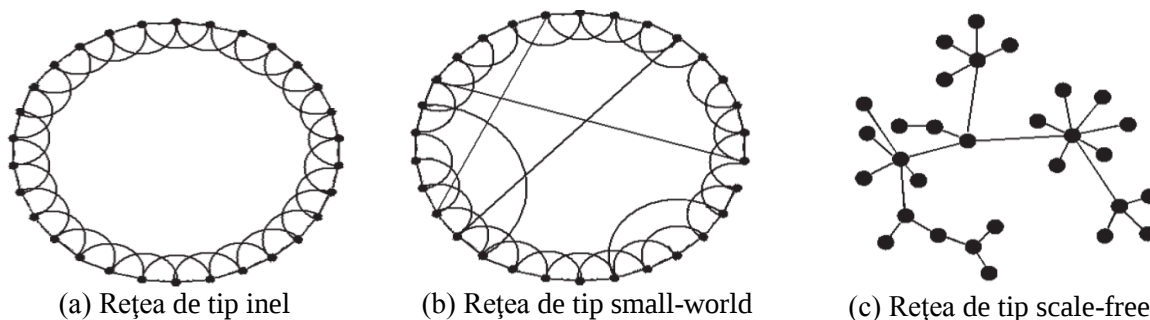


Figura 3.1: Exemple de rețele generate folosind modelul WS - (a) și (b), respectiv modelul BA - (c) (luate de la [24]).

CAPITOLUL 4

Propagarea Opiniilor în Rețelele de Socializare Online

4.1. Modelarea Propagării Opiniilor în Rețelele de Socializare Online

În general, modelele de propagare de opinii în rețelele de socializare online pot fi clasificate în două categorii: modele bazate pe grafuri nedirecționate pentru detectarea propagării opiniilor și modele bazate pe grafuri direcționate pentru detectarea propagării opiniilor. Principala caracteristică distinctivă a acestor două categorii de modele este dată de tipurile de relații stabilite între membrii rețelelor de socializare online. Astfel, în modelele bazate pe grafuri nedirecționate, relațiile sunt simetrice, în timp ce în modelele bazate pe grafuri direcționate, relațiile sunt asimetrice.

Modelele bazate pe grafuri nedirecționate pentru detectarea propagării de opinii pot fi clasificate, în funcție de modul de reprezentare a opiniilor, astfel:

- modele de opinii binare. Modelele de opinii binare constituie un caz particular al modelelor de opinii discrete. În modelele de opinii binare, spațiul în care opiniile pot lua valori este redus la numai două valori. Concret, fiecare individ $u \in V$ are în orice moment de timp t , o singură opinie $o(u, t)$ care poate lua una dintre cele două valori posibile, de multe ori o valoare binară de 0 sau 1. Interpretăm opinia binară pe care un individ o are prin faptul că acceptă sau nu un anumit lucru. Ca exemplu din lumea reală în care opiniile indivizilor sunt valori binare, putem menționa situația în care oamenii apreciază sau nu un produs sau un candidat în alegerile electorale. Un spațiu de opinii binare este folosit în modelele de propagare de opinii, precum modelul Voter [5].
- modele de opinii continue. Aceste modele se caracterizează prin faptul că opinia $o(u, t)$ a fiecărui individ $u \in V$ este exprimată printr-un număr real, de mai multe ori în intervalul $[0, 1]$. Pentru a înțelege utilitatea opiniilor modelate prin valori continue într-un interval, trebuie să ținem cont de faptul că o persoană poate avea opinii de intensități diferite despre aceeași entitate țintă de-a lungul timpului. Ca modele de propagare de opinii care folosesc opinii continue amintim: modelul Sznajd [6], modelul Deffuant [7] sau modelul Hegselmann-Krause [8].
- modele cu vectori de opinii continue. În aceste modele [25], [26] fiecare individ $u \in V$ are o opinie modelată printr-o valoare continuă despre fiecare entitate dintr-o mulțime de entități țintă partajate de către toate persoanele dintr-o rețea de socializare online.

Modelul Independent Cascade (IC) [9] și modelul Linear Threshold (LT) [9] sunt două modele de propagare de opinii binare bazate pe grafuri direcționate. Aceste modele au fost propuse inițial pentru procesul propagării informațiilor în rețelele de socializare online, dar ele pot fi aplicate, de asemenea, pentru procesul propagării opiniilor.

4.2. Discuții

Propagarea opiniilor și difuzarea informațiilor în rețelele de socializare online nu sunt similare. Există mai multe diferențe între aceste procese. Propagarea opiniilor este un fenomen mult mai complex, în care influențatorii au scopul, de cele mai multe ori disimulat, de a convinge alte persoane să își schimbe opiniile. În schimb, difuzarea informațiilor se referă la comunicarea unor anumite fapte fără intenția de a schimba opiniile altor persoane. De asemenea, în timp ce informația este verificată, opiniile nu sunt în mod necesar bazate pe dovezi. În ceea ce privește zvonurile, acestea sunt tendențioase și ascund adesea intenția de a face rău.

Toate modelele enumerate în secțiunea 4.1 folosesc valori numerice pentru opiniile indivizilor. Într-o analiză a rețelelor reale de socializare online, este necesar să se extragă opiniile din mesaje schimbate între indivizi. Mai mult decât atât, în toate aceste modele se presupune că fiecare individ are o opinie inițială. În realitate nu toți indivizii au o opinie, dar ei și-o formează de-a lungul timpului. Formarea de opinie într-o rețea de socializare online este rezultatul interacțiunilor cu alte persoane care au o opinie.

Propagarea opiniilor ar trebui să fie studiată în relație cu influența dintre indivizi deoarece influența socială reprezintă un factor decisiv în declanșarea răspândirii opiniilor. Unele modele prezentate (de exemplu, modelul Sznajd, modelul Deffuant, modelul Hegselmann-Krause și modelul Linear Threshold) au studiat propagarea opiniilor aplicând diferite teorii sociale. Ideea comună ce fundamentează aceste patru modele este că indivizii interacționează numai dacă au opinii similare. În modelul Voter, nu există niciun parametru pentru a indica nivelul de influență între doi indivizi, aceștia interacționând în mod aleatoriu. În schimb, în modelul Independent Cascade, nivelul de influență între doi indivizi este specificat de parametrul $p_{v,u}$.

Modelele bazate pe grafuri nedirecționate pentru detectarea propagării opiniilor au fost simulate pe diverse topologii de rețea: rețele de tip scale-free sau rețele de tip small-world. Rețelele de tip small-world nu țin seama de faptul că persoane noi se pot alătura rețelei de-a lungul timpului. Din punct de vedere al propagării opiniilor, persoanele noi care devin membri într-o rețea de socializare online pot juca un rol cheie. Unele dintre aceste persoane pot fi mult mai influențatoare în rețea decât membrii existenți. În plus, persoanele noi pot stabili legături directe strategice cu unii membri ai rețelei, în scopul de a-i influența.

CAPITOLUL 5

Provocări în Cercetarea Mineritului Opiniilor dintr-o Perspectivă a Prelucrării Limbajului Natural

5.1. Preprocesarea de Texte

De regulă, preprocesarea de texte este o parte componentă a algoritmilor de extragere de opinii și de aceea rezultatele acestui proces trebuie să fie precise. În această subsecțiune, vom descrie două sarcini de preprocesare aplicabile unui text ce conține opinii: tokenizarea și etichetarea cu părți de vorbire. Tokenizarea este definită de Manning și Schütze [27, p. 124] ca un proces de divizare a textului de intrare în unități (numite *token-uri*) cu o semnificație independentă cum ar fi: cuvinte, numere, semne de punctuație.

Etichetarea cu părți de vorbire este procesul de etichetare gramaticală cu partea de vorbire corespunzătoare a tuturor cuvintelor dintr-un text.

Etichetele folosite în acest proces pot face referire la părți de vorbire la modul general (de exemplu, substantive, verbe, adjective etc.), dar pot furniza și informații suplimentare cu privire la unele caracteristici morfologice ale cuvintelor: număr, gen, persoană, timp sau aspect.

În cazul procesului de etichetare cu părți de vorbire, ambiguitățile apar din cauza faptului că un cuvânt poate avea mai multe valori morfologice în funcție de context. Această problemă de alegere a părții de vorbire adecvate poate fi rezolvată prin luarea în considerare a caracteristicilor morfologice ale cuvintelor. Odată ce o parte de vorbire a fost identificată, aceasta oferă informații despre posibilele părți de vorbire care pot apărea în aceeași propoziție.

5.2. Extragerea Cuvintelor de Opinie

Pe baza unor analize morfologice și sintactice, putem stabili anumite relații de dependență între cuvintele dintr-o propoziție. Fiecare relație de dependență este o relație gramaticală binară, adică între două cuvinte dintre care unul este *guvernator* și celălalt *dependent* [28].

Marneffe ș.a. [28] au dezvoltat o ierarhie de relații de dependență între cuvintele dintr-o propoziție. O relație de dependență de tip Stanford este reprezentată în forma prescurtată *nume_relație(guvernator, dependent)*, unde dependența există între cuvântul *guvernator* și cuvântul *dependent*. O descriere exhaustivă a mulțimii de relații de dependență de tip Stanford poate fi găsită în lucrarea sursă [28]. Din punct de vedere al mineritului opiniilor, relațiile de dependență între cuvintele dintr-o propoziție sunt importante deoarece există unele relații de dependență care pot fi folosite pentru a extrage opiniile cu privire la entitățile țintă. Concret, aceste dependențe reprezintă relații binare între cuvintele de opinie și entitățile țintă.

Problema rezolvării coreferințelor presupune identificarea „tuturor expresiilor substantive care se referă la aceeași entitate într-un text” [29]. Studiată intens în PLN, această problemă trebuie luată în considerare în procesul mineritului opiniilor deoarece, în caz contrar, fie „opiniile vor fi pierdute”, fie „opiniile vor fi atribuite în mod eronat altor entități” [30].

5.3. Detectarea Polarității Cuvintelor de Opinie

Modificatorii de valență [31] sunt cuvinte sau expresii care, atunci când sunt folosite într-o propoziție, intensifică, diminuează sau schimbă de la pozitiv la negativ sau de la negativ la pozitiv orientarea semantică a cuvintelor de opinie din acea propoziție. De exemplu, în propoziția „Nu-mi place zăpada.”, valența pozitivă pentru verbul „place” este schimbată într-o valență negativă din cauza folosirii negației „nu”. Putem distinge mai multe categorii de modificatori de valență: negațiile, ironiile, verbele modale etc.

Cuvintele de opinie pot exprima sentimente diferite în funcție de contextul în care aceste cuvinte de opinie apar. Concret, pot fi situații în care un cuvânt de opinie exprimă, într-un anumit context, un sentiment pozitiv, dar are polaritatea neutră sau negativă în afara acestui context. Explicația acestui fapt este următoarea: unele cuvinte, numite *cuvinte ambigue*, pot avea mai multe sensuri, adică sunt *cuvinte polisemantice*. De exemplu, în expresiile „preț mare” și „rezoluție mare”, adjectivul „mare” are un sentiment negativ în prima expresie și un sentiment pozitiv în cea de a doua expresie [32]. Pentru a determina sentimentul cuvintelor de opinie, utilizarea exclusivă a vocabularelor de opinii nu este suficientă deoarece aceste vocabulare nu permit identificarea exactă a sentimentului contextual exprimat de un cuvânt de opinie. Metode specifice ar trebui create pentru a identifica în mod corect sentimentul exprimat de cuvintele de opinie într-un context dat [32].

Dezambiguizarea sensurilor cuvintelor de opinie este un domeniu situat la intersecția dintre dezambiguizarea sensurilor cuvintelor și analiza sentimentelor. Dezambiguizarea sensurilor cuvintelor este definită în cartea lui Manning și Schütze [27, p. 229] ca fiind procesul în care determinăm care dintre sensurile unui cuvânt ambiguu este folosit într-un context dat. Dezambiguizarea sensurilor cuvintelor presupune identificarea sensului contextual al cuvintelor ambigue care exprimă atât opinii, cât și fapte. În dezambiguizarea sensurilor cuvintelor de opinie suntem interesați să identificăm, mai întâi, contextul în care este folosit cuvântul de opinie, iar apoi, sentimentul exprimat de acesta.

5.4. Rolul Etichetării cu Părți de Vorbire a Cuvintelor în Mineritul Opiniilor

În această secțiune, vom descrie implementarea unui etichetator cu părți de vorbire a cuvintelor dintr-o propoziție folosind modelul Markov ascuns bazat pe trigrame, în conformitate cu descrierea propusă de Brants [33]. Abordarea noastră este una experimentală. Testele sunt făcute cu scopul de a obține rezultate în funcție de diverși parametri. Primele experimente pun în evidență modul în care precizia etichetatorului variază atunci când dimensiunea corpusului de date de antrenare se modifică.

În zilele noastre, când Internetul a devenit un fenomen larg răspândit, mediile colaborative online, cum ar fi chaturile, forumurile, blogurile sau wiki-urile, au devenit mijloace importante de comunicare, dar și de influență. Următoarele experimente analizează rezultatele obținute atunci când etichetăm cu partea de vorbire corespunzătoare cuvinte atât din textele scrise din mediile colaborative online, cât și din textele literare. Această clasificare ne permite să arătăm modul în care precizia etichetatorului se schimbă atunci când este antrenat și testat pe texte din domenii diferite. Pentru a face testele, vom folosi corpusul Brown [34], un corpus de wiki, un corpus de forum și un corpus de chat.

CAPITOLUL 6

Analiza Propagării Opiniilor la Nivel de Utilizator în Firele de Discuție din Forumuri

6.1. Introducere

Până în prezent, problema propagării opiniilor a fost studiată în rețelele de socializare online. Studiile anterioare au pus accent pe analiza propagării opiniilor în rețele generate folosind modele de rețea de tip Watts-Strogatz [22] și de tip Barabási-Albert [23]. Rețelele obținute folosind cele două tipuri de modele au proprietăți similare cu cele ale rețelelor de socializare online [20, 21]. În aceste studii, interacțiunile dintre indivizii vecini sunt stabilite în mod aleatoriu. De asemenea, opiniile sunt repartizate, inițial, în mod aleatoriu indivizilor.

În acest capitol, propunem o analiză a propagării opiniilor în alte tipuri de medii sociale decât rețelele de socializare online, adică în firele de discuție din forumuri. Metoda noastră constă în a determina dacă participanții la discuție ajung la un acord total sau parțial, ori continuă să aibă opinii diferite sau chiar contrare. Soluția noastră se bazează pe opiniile exprimate de utilizatori despre diferiți termeni substantivali legați semantic de subiectul discuției. În acest scop, vom extrage opiniile utilizatorilor din postările trimise de ei. Metoda propusă este evaluată prin urmărirea propagării opiniilor în fire de discuții reale din forumuri.

6.2. Analiza Firelor de Discuție din Forumuri

Această secțiune oferă o analiză a firelor de discuție din forumuri pentru a înțelege obiectul cercetării noastre în acest capitol. Formal, un forum online este compus dintr-o mulțime de fire de discuție definite după cum urmează:

Definiția 6.1 (Fir de Discuție). *Fiind dat un forum de discuții online, definim un fir de discuție la momentul de timp t_τ , $\tau \in \mathbb{N}^*$, printr-un cvintuplu $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$, unde $t_1, t_2, \dots, t_\tau, \dots$ sunt momente de timp discrete pe axa timpului, ce corespund acțiunilor, de tipul adăugării de noi postări, care au loc într-un fir de discuție de-a lungul timpului.*

Un fir de discuție este caracterizat, în mod special, de un subiect S_{DT} și, în mod general, de o topică T_{DT} . Astfel, fiecare fir de discuție este despre un anumit subiect S_{DT} , iar mai multe fire de discuție sunt grupate sub aceeași topică T_{DT} . Postările $p \in P_{DT}(t_\tau)$ reprezintă intervenții ale utilizatorilor $u \in U_{DT}(t_\tau)$ despre subiectul S_{DT} în respectivul fir de discuție, până la momentul de timp t_τ .

Fiecare postare $p \in P_{DT}(t_\tau)$ este un răspuns la o altă postare. La începutul firului de discuție, care corespunde momentului de timp t_1 , un utilizator trimite o postare, iar apoi, în momentele de timp t_τ , $\tau \in \mathbb{N}$, $\tau \geq 2$, alți utilizatori intervin în discuție. În rețelele de socializare online, precum Facebook, fiecare utilizator interacționează, de regulă, numai cu utilizatorii cu care a stabilit în viața reală relații de prietenie. În schimb, într-un fir de discuție dintr-un forum, un utilizator $u \in U_{DT}(t_\tau)$ interacționează, la momentul de timp t_τ , cu orice alt utilizator care a participat la discuție până la momentul t_τ , chiar dacă ei nu se cunosc, ceea ce înseamnă că utilizatorul $u \in U_{DT}(t_\tau)$ poate

să răspundă la postările anterioare pe care ceilalți utilizatori le-au scris. Cu alte cuvinte, toate postările trimise de către toți utilizatorii sunt vizibile oricărui participant la discuție.

Într-un fir de discuție, la momentul de timp t_τ , două tipuri de relații pot fi definite: $R_{UP}(t_\tau)$ și $R_{PP}(t_\tau)$, $R_{DT}(t_\tau) = R_{UP}(t_\tau) \cup R_{PP}(t_\tau)$. O relație binară parțială $R_{UP}(t_\tau) \subset U_{DT}(t_\tau) \times P_{DT}(t_\tau)$ este definită între mulțimile $U_{DT}(t_\tau)$ și $P_{DT}(t_\tau)$ astfel încât, dacă $(u, p) \in R_{UP}(t_\tau)$, postarea $p \in P_{DT}(t_\tau)$ a fost scrisă de către utilizatorul $u \in U_{DT}(t_\tau)$. O relație binară parțială $R_{PP}(t_\tau) \subset P_{DT}(t_\tau) \times P_{DT}(t_\tau)$ se stabilește între postările din mulțimea $P_{DT}(t_\tau)$ astfel încât, dacă $(p_1, p_2) \in R_{PP}(t_\tau)$, postarea $p_1 \in P_{DT}(t_\tau)$ răspunde la o altă postare $p_2 \in P_{DT}(t_\tau)$.

Luând în considerare toate explicațiile de mai sus, pentru două secvențe $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ și $(T_{DT}, S_{DT}, U_{DT}(t_{\tau'}), P_{DT}(t_{\tau'}), R_{DT}(t_{\tau'}))$ ale aceluiași fir de discuție, corespunzătoare momentelor de timp t_τ și $t_{\tau'}$, $t_{\tau'} > t_\tau$, putem scrie următoarele relații: $U_{DT}(t_\tau) \subseteq U_{DT}(t_{\tau'})$, $P_{DT}(t_\tau) \subset P_{DT}(t_{\tau'})$, $R_{UP}(t_\tau) \subset R_{UP}(t_{\tau'})$ și $R_{PP}(t_\tau) \subset R_{PP}(t_{\tau'})$.

Relațiile stabilite între utilizatorii dintr-o rețea de socializare online ne permit să reprezentăm comunitatea de utilizatori printr-un graf social $G_s = (V_s, E_s)$, unde V_s este mulțimea utilizatorilor, iar E_s este mulțimea relațiilor care conectează utilizatorii. Spre deosebire de o rețea de socializare online, un fir de discuție dintr-un forum este modelat pe o structură de graf orientat, folosind relațiile R_{PP} .

6.3. Formularea Problemei

Înainte de a defini problema propagării opiniilor la nivel de utilizator în firele de discuție din forumuri, vom introduce câteva definiții și notații necesare.

Un fir de discuție dintr-un forum cuprinde o serie de postări trimise de utilizatorii $u \in U_{DT}(t_\tau)$ despre același subiect S_{DT} . Subiectul S_{DT} permite utilizatorilor care participă la discuție să își exprime opiniile cu privire la diferite idei exprimate prin termeni substantivali. Mai mult decât atât, acești termeni substantivali sunt legați în mod semantic de unul dintre termenii substantivali care apar în denumirea subiectului S_{DT} . Prin urmare, putem scrie următoarea definiție:

Definiția 6.2 (Vocabular de Termeni Substantivali). *Într-un fir de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$, toți termenii substantivali despre care utilizatorii $u \in U_{DT}(t_\tau)$ și-au exprimat opiniile și care sunt semantic legați de unul dintre termenii substantivali care apar în denumirea subiectului S_{DT} definesc un vocabular de termeni substantivali. Folosim notația $V^d_{DT}(t_\tau) = \{s_1, s_2, \dots, s_d\}$ pentru a indica un vocabular d -dimensional de termeni substantivali partajați de toți utilizatorii din firul de discuție din forum, până la momentul de timp t_τ , $\tau \in \mathbb{N}^*$.*

Mărimea și structura vocabularului de termeni substantivali poate varia în timp. Postările noi care sunt adăugate firului de discuție pot conține opinii cu privire la noi termeni substantivali. În cele ce urmează, vom face presupunerea că opiniile utilizatorilor sunt modelate prin valori numerice.

Definiția 6.3 (Spațiu de Opinii Discrete). *Fiind date la timpul t_τ , $\tau \in \mathbb{N}^*$, firul de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum și vocabularul corespunzător $V^d_{DT}(t_\tau)$ de termeni substantivali, atunci opiniile fiecărui utilizator $u \in U_{DT}(t_\tau)$ despre termenii substantivali din vocabularul $V^d_{DT}(t_\tau)$ pot fi reprezentate printr-un vector într-un spațiu d -dimensional $OS^d_{DT} = \{-1, 0, +1\}^d$ de opinii discrete.*

În acest spațiu, fiecare axă corespunde unui termen substantival, iar valorile de pe fiecare axă corespund unor posibile opinii ale utilizatorilor despre termenul substantival respectiv. În cazul studiat de noi, opiniile sunt valori discrete din mulțimea $\{-1, 0, +1\}$.

Se consideră un vector $o_{DT}^d(t_\tau) = [o_1 \ o_2 \ \dots \ o_d]^T$ la momentul de timp t_τ , $\tau \in \mathbb{N}^*$, în spațiul d -dimensional OS_{DT}^d de opinii discrete și care corespunde utilizatorului $u \in U_{DT}(t_\tau)$. Fiecare intrare o_k în vectorul $o_{DT}^d(t_\tau)$ reprezintă opinia utilizatorului $u \in U_{DT}(t_\tau)$ despre fiecare termen substantival s_k din vocabularul $V_{DT}^d(t_\tau)$, până la momentul de timp t_τ , $\tau \in \mathbb{N}^*$.

Intrările de opinii o_k din vectorul $o_{DT}^d(t_\tau)$ pot lua una dintre următoarele valori: -1, 0, sau 1, unde valoarea -1 indică o opinie negativă, valoarea 0 indică o opinie neutră, iar valoarea +1 indică o opinie pozitivă. Dacă utilizatorul $u \in U_{DT}(t_\tau)$ nu și-a exprimat opinia cu privire la termenul substantival $s_k \in V_{DT}^d(t_\tau)$ până la momentul de timp t_τ , $\tau \in \mathbb{N}^*$, atunci considerăm valoarea 0 pentru intrarea $o_k \in o_{DT}^d(t_\tau)$. Dacă până la momentul de timp t_τ , $\tau \in \mathbb{N}^*$, utilizatorul $u \in U_{DT}(t_\tau)$ și-a exprimat mai multe opinii despre același termen substantival, numai ultima sa opinie este luată în considerare.

Inspirați de cercetarea lui Stavrianou [35], definim reprezentarea unui fir de discuție dintr-un forum, utilizând o structură de date de tipul unui graf orientat, care ține cont de opiniile exprimate de către utilizatori în postările lor, după cum urmează:

Definiția 6.4 Fiind dat la timpul t_τ , $\tau \in \mathbb{N}^*$, firul de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum, asociem firului de discuție un graf orientat $G_{DT}(t_\tau)(V_{DT}(t_\tau), E_{DT}(t_\tau))$, unde:

- $V_{DT}(t_\tau) = \{v_j \mid v_j = (v_j^p, v_j^u, v_j^{tm}, v_j^{op}), v_j^p \in P_{DT}(t_\tau), v_j^u \in U_{DT}(t_\tau), v_j^{tm} \in \mathbb{N}, v_j^{op} = [o_1 \ o_2 \ \dots \ o_m]^T, m \leq d, v_j^{op} \subset OS_{DT}^m, j \in [1, n]\}$ reprezintă mulțimea nodurilor în graful $G_{DT}(t_\tau)$, iar n reprezintă numărul nodurilor în graful $G_{DT}(t_\tau)$;
- $E_{DT}(t_\tau) = \{e_1, e_2, \dots, e_{n-1}\}$ reprezintă mulțimea arcelor astfel încât, dacă $e = (v_i, v_j) \in E_{DT}(t_\tau)$, atunci postarea v_i este un răspuns la postarea v_j .

În graful $G_{DT}(t_\tau)$, fiecare nod $v_j \in V_{DT}(t_\tau)$ corespunde unei postări v_j^p scrise de către un utilizator v_j^u la momentul de timp v_j^{tm} . Postarea v_j^p exprimă opinii în vectorul m -dimensional v_j^{op} de opinii, unde m este egal cu sau mai mic decât dimensiunea d a vocabularului $V_{DT}^d(t_\tau)$ de termeni substantivali. Să considerăm încă două definiții, după cum urmează:

Definiția 6.5 (Matricea de Opinii Termeni-Utilizatori). Fiind date la momentul de timp t_τ , $\tau \in \mathbb{N}^*$, firul de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum, graful $G_{DT}(t_\tau)(V_{DT}(t_\tau), E_{DT}(t_\tau))$ asociat firului de discuție și vocabularul d -dimensional $V_{DT}^d(t_\tau)$ de termeni substantivali, atunci construim matricea de opinii termeni-utilizatori $A_{T-U}(t_\tau) = [A_1(t_\tau) \ A_2(t_\tau) \ \dots \ A_n(t_\tau)]$ de dimensiune $d \times n$, unde n reprezintă cardinalul mulțimii de utilizatori $U_{DT}(t_\tau)$. Fiecare coloană $A_u(t_\tau) = [a_{1,u}(t_\tau) \ a_{2,u}(t_\tau) \ \dots \ a_{d,u}(t_\tau)]^T$ corespunde unui utilizator $u \in U_{DT}(t_\tau)$ și reprezintă vectorul d -dimensional de opinii al utilizatorului u în spațiul de opinii discrete OS_{DT}^d .

Definiția 6.6 (Matricea de Similaritate Utilizator-Utilizator). Fiind date la momentul de timp t_τ , $\tau \in \mathbb{N}^*$, firul de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum, graful $G_{DT}(t_\tau)(V_{DT}(t_\tau), E_{DT}(t_\tau))$ asociat firului de discuție, vocabularul d -dimensional $V_{DT}^d(t_\tau)$ de termeni substantivali, matricea de opinii termeni-utilizatori $A_{T-U}(t_\tau)$ de dimensiune $d \times n$ și o măsură de similaritate între

doi vectori, atunci putem construi o matrice $B_{U-U}(t_\tau)$ de similaritate între utilizatori. Considerăm că dimensiunea matricei $B_{U-U}(t_\tau)$ este $n \times n$. O intrare $b_{k,h}(t_\tau)$ în matricea $B_{U-U}(t_\tau)$, corespunzătoare liniei k și coloanei h , reprezintă similaritatea între utilizatorii $k, u \in U_{DT}(t_\tau)$, din perspectiva opiniilor exprimate de acești utilizatori.

Concret, intrarea $b_{k,h}(t_\tau)$ măsoară similaritatea între vectorii coloană $A_k(t_\tau) = [a_{1,k}(t_\tau) \ a_{2,k}(t_\tau) \ \dots \ a_{d,k}(t_\tau)]^T$ și $A_h(t_\tau) = [a_{1,h}(t_\tau) \ a_{2,h}(t_\tau) \ \dots \ a_{d,h}(t_\tau)]^T$ ai matricei $A_{T-U}(t_\tau)$. Calculăm similaritatea dintre doi vectori utilizând cosinusul unghiului format între aceștia, după cum urmează:

$$sim_{cos}(A_k(t_\tau), A_h(t_\tau)) = \frac{\sum_{i=1}^d a_{i,k}(t_\tau) a_{i,h}(t_\tau)}{\sqrt{\sum_{i=1}^d (a_{i,k}(t_\tau))^2} \sqrt{\sum_{i=1}^d (a_{i,h}(t_\tau))^2}} \quad (6.1)$$

Măsura cosinus returnează un scor între -1 și +1. În acest caz, valoarea -1 înseamnă dezacord total între utilizatori (utilizatorii au opinii cu sentimente opuse despre aceiași termeni substantivali), în timp ce valoarea +1 înseamnă acord total între utilizatori (utilizatorii au opinii cu sentimente identice despre aceiași termeni substantivali).

În cele din urmă, putem să definim în mod formal problema propagării opiniilor la nivel de utilizator într-un fir de discuție din forum, prin extinderea definiției propuse de Zafarani ș.a. [36]:

Definiția 6.7 (Problema Propagării Opiniilor la Nivel de Utilizator într-un Fir de Discuție). *Fiind date la momentul de timp t_τ , $\tau \in \mathbb{N}^*$, firul de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum, graful $G_{DT}(t_\tau)(V_{DT}(t_\tau), E_{DT}(t_\tau))$ asociat firului de discuție, vocabularul d -dimensional $V_{DT}^d(t_\tau)$ de termeni substantivali, matricea $A_{T-U}(t_\tau)$ de opinii termeni-utilizatori, matricea $B_{U-U}(t_\tau) = \{b_{k,h}(t_\tau)\}$ de similaritate utilizator-utilizator și o submulțime de utilizatori $U'_{DT}(t_\tau) \subset U_{DT}(t_\tau)$ care au vectorii de opinii similari la orice moment de timp t_i , $t_i \geq t_\tau$, adică $b_{u_2 u_1} \leq \varepsilon_1$, $\forall u_2, u_1 \in U'_{DT}(t_\tau)$, și care inițiază propagarea de opinii la momentul de timp t_τ , atunci un utilizator $u \in U_{DT}(t_j) \setminus U'_{DT}(t_\tau)$ este influențat la momentul de timp t_j numai dacă este îndeplinită condiția de mai jos:*

$$b_{u_3, u}(t_j) \leq b_{u_1, u}(t_j) + \varepsilon_2 \quad (6.2)$$

unde $u_3 \in U_{DT}(t_j) \setminus U'_{DT}(t_\tau)$, $\forall u_1 \in U'_{DT}(t_\tau)$ și parametrii $\varepsilon_1, \varepsilon_2$ pot fi euristice setați.

6.4. Arhitectura Abordării Propuse

Sistemul care implementează metoda propusă de noi este format din mai multe module, grupate după cum se arată în Figura 6.1. În primul rând, folosind instrumentul Stanford CoreNLP¹, sarcini specifice de prelucrare a limbajului natural sunt aplicate fiecărei postări din firul de discuție [27]: tokenizarea, etichetarea cu părți de vorbire a cuvintelor, parsarea de dependențe, rezolvarea coreferințelor.

¹ <http://nlp.stanford.edu/software/corenlp.shtml>

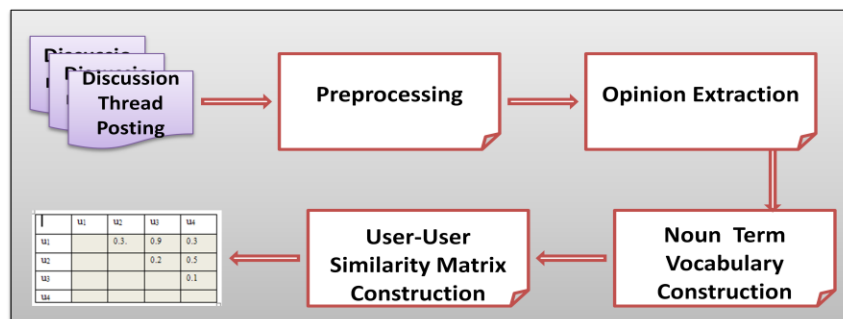


Figura 6.1: Arhitectura generală a problemei propagării opiniilor la nivel de utilizator în firele de discuție din forumuri.

În cele ce urmează, vom considera că toate cuvintele de opinie pot fi folosite ca părți de vorbire precum adjective, adverbe și verbe. În scopul de a identifica cuvintele de opinie din postări, au fost utilizate relațiile de dependență Stanford [37]: „dobj”, „nsubj”, „amod”, „acomp”, „advmod”, „xcomp” și „neg”. Din punct de vedere al mineritului opiniilor, aceste relații de dependență indică relații între termenii substantivali și cuvintele care conțin opinii despre aceștia.

Pentru a identifica sentimentul cuvintelor de opinie, am selectat patru vocabulare de opinii, care vor fi folosite în experimentele noastre: SentiWordNet [1], Micro-WNOp [2], vocabularul de subiectivitate MPQA [3] și vocabularul de opinii creat de Bing Liu [4]. În problema propagării opiniilor la nivel de utilizator pe care o studiem, opiniile utilizatorilor sunt modelate prin valori numerice. După extragerea cuvintelor de opinie din postări, este necesară maparea fiecărui cuvânt de opinie la o valoare numerică din mulțimea $\{-1, 0, +1\}$, unde o valoare de -1 corespunde unui cuvânt de opinie de sentiment negativ, o valoare de 0 corespunde unui cuvânt de opinie de sentiment neutru, iar o valoare de +1 corespunde unui cuvânt de opinie de sentiment pozitiv. Pentru fiecare vocabular de opinii, noi propunem algoritmi specifici astfel încât să asociem valori numerice cuvintelor de opinie. În cele din urmă, vom construi vectorul de opinii pentru fiecare participant la discuție, precum și matricea de similaritate utilizator-utilizator B_{U-U} .

6.5. Experimente

Metoda propusă în acest capitol este evaluată folosind firele de discuție din corpusul creat de Walker ș.a. [38]. Acest corpus conține fire de discuție de pe forumul 4Forums¹. Corpusul este disponibil gratuit, iar informațiile despre fiecare fir de discuție pot fi găsite în formatul JSON².

Firele de discuție de pe forumul 4Forums au caracteristici de structură, ce au fost prezentate în secțiunea 6.2. Mai mult decât atât, datele conținute de fiecare fir de discuție din acest corpus ne permit să construim graful corespunzător. Astfel, fiecare fir de discuție cuprinde o mulțime de postări, iar fiecare postare conține mai multe componente, după cum urmează: autorul postării, textul postării, momentul de timp când a fost trimisă pe firul de discuție și postarea părinte.

Fără a pierde din vedere generalitatea problemei formulate în Definiția 6.7, în cele ce urmează vom studia cazul în care $t_r = t_l$ și mulțimea $U'_{DT}(t_l) = \{u_l\}$. Concret, considerăm că procesul de propagare de opinii este inițiat, la momentul de timp t_l , de către primul utilizator (denumit u_l)

¹ <http://www.4forums.com>

² <http://json.org/>

care a trimis o postare în respectivul fir de discuție. În acest caz, indivizii țintă ai procesului de propagare de opinii vor fi toți utilizatorii care intervin în discuție de-a lungul timpului (la momentele de timp $t_j > t_1$). În cazul în care, la momentul de timp $t_j > t_\tau$, ecuația $1 - \varepsilon \leq b_{u,u_1}(t_j) \leq 1$ se verifică pentru toți utilizatorii din mulțimea $U_{DT}(t_j) \setminus \{u_1\}$, atunci procesul de propagare de opinii este caracterizat de un acord total între utilizatorii care au participat la discuție.

Din corpusul creat de Walker ș.a. [38] am selectat două fire de discuție pe subiectul denumit „Existența lui Dumnezeu”. Considerăm structura celor două fire de discuție la trei momente de timp diferite. Tabelul 6.1 arată, pentru fiecare fir de discuție în parte, cele trei momente de timp t_j și numărul de utilizatori $U_{DT}(t_j)$ care au trimis postări până la momentul de timp t_j .

Tabel 6.1. Statistici cu privire la firele de discuție din forum

Identificatorul Firului de Discuție dintr-un Forum	Moment de Timp Discret (t_j)	Număr de Utilizatori ($U_{DT}(t_j)$)
Fir de Discuție 1	t_{49}	7
	t_{100}	14
	t_{149}	15
Fir de Discuție 2	t_{60}	18
	t_{121}	19
	t_{192}	20

În Figura 6.2, reprezentăm similaritățile dintre vectorul de opinii corespunzător utilizatorului u_1 și vectorii de opinii corespunzători celorlalți utilizatori care intervin în discuție după momentul de timp t_1 . Concret, aceste valori de similaritate corespund primei linii din matricea de similaritate utilizator-utilizator B_{U-U} . Valorile de similaritate sunt calculate pentru ambele fire de discuție și pentru fiecare dintre cele patru vocabulare de opinii.

În Figura 6.2, utilizatorii din mulțimea $U_{DT}(t_j) \setminus \{u_1\}$ sunt reprezentați pe axa X. În Figura 6.2 (a-c), se poate observa că similaritatea vectorilor de opinii corespunzători utilizatorilor u_1 și u_5 scade de-a lungul timpului. În Figura 6.2 (d-f), se poate observa că similaritatea vectorilor de opinii corespunzători utilizatorilor u_1 și u_4 scade de-a lungul timpului, iar similaritatea vectorilor de opinii corespunzători utilizatorilor u_1 și u_{12} crește în timp.

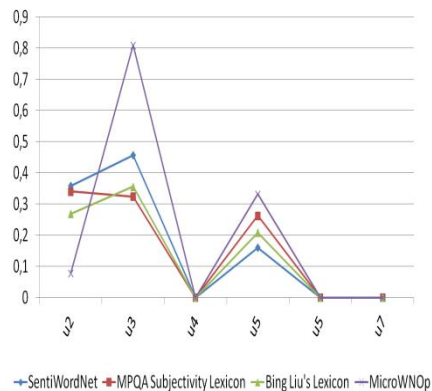
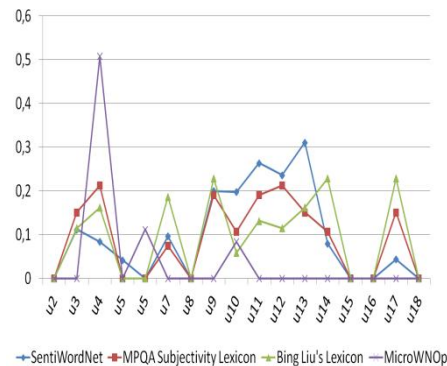
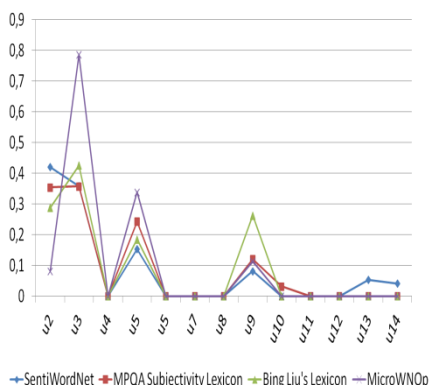
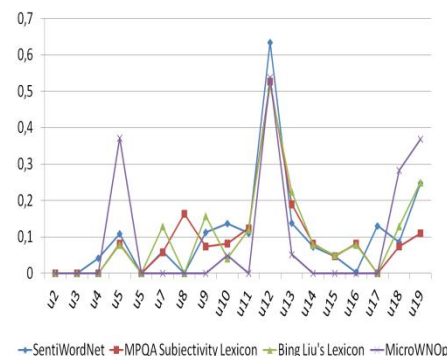
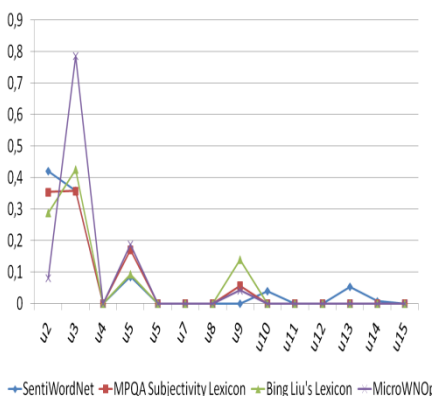
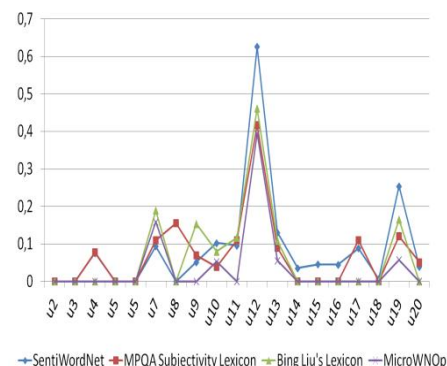
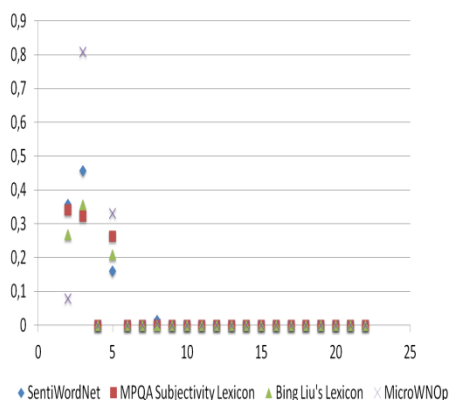

 (a) Fir de Discuție 1 – t_{49}

 (d) Fir de Discuție 2 - t_{60}

 (b) Fir de Discuție 1 - t_{100}

 (e) Fir de Discuție 2 - t_{121}

 (c) Fir de Discuție 1 - t_{149}

 (f) Fir de Discuție 2 - t_{192}

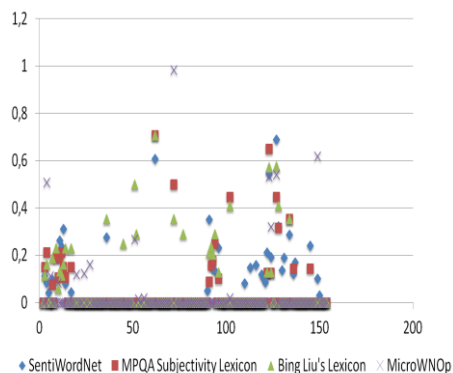
Figura 6.2: Similarități între vectorul de opinii corespunzător utilizatorului u_1 și vectorii de opinii corespunzători celorlalți utilizatori care se alătură discuției până la momentul de timp t_r în firul de discuție 1 (a-c), respectiv în firul de discuție 2 (d-f).

În Figura 6.3, reprezentăm grafic toate valorile de similaritate din matricea de similaritate utilizator-utilizator B_{U-U} , pentru ambele fire de discuție și pentru fiecare dintre cele patru vocabulare de opinii. În Figura 6.3 (a-c) pentru firul de discuție 1, se poate observa că în matricea B_{U-U} există de-a lungul timpului mai multe valori mai mari decât zero. În schimb, în Figura 6.3

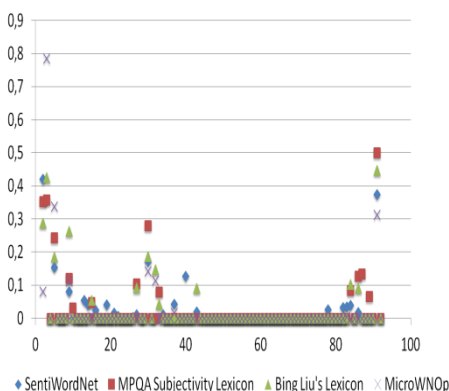
(d-f) pentru firul de discuție 2, putem constata că de-a lungul timpului există mai puține valori mai mari decât zero în matricea B_{U-U} .



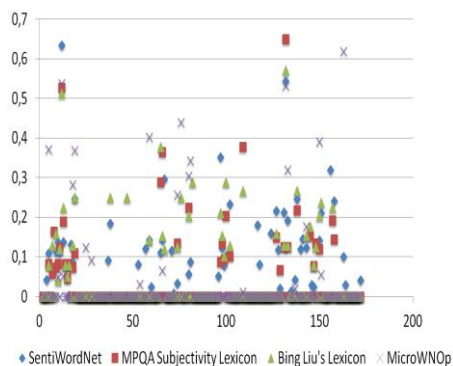
(a) Fir de Discuție 1 – t_{49}



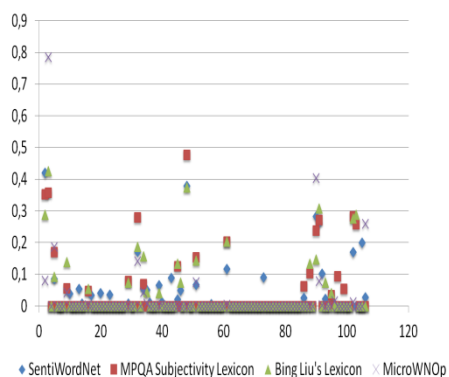
(d) Fir de Discuție 2 – t_{60}



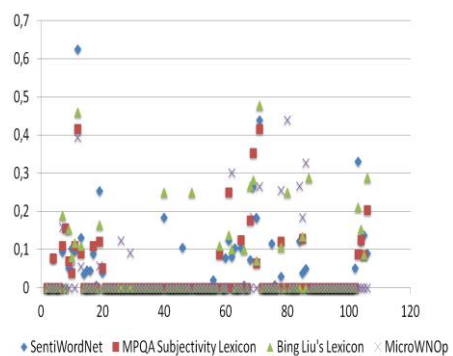
(b) Fir de Discuție 1 – t_{100}



(e) Fir de Discuție 2 – t_{121}



(c) Fir de Discuție 1 – t_{149}



(f) Fir de Discuție 2 – t_{192}

Figura 6.3 Valorile matricei de similaritate utilizator-utilizator B_{U-U} până la momentul de timp t_τ în firul de discuție 1 (a-c), respectiv în firul de discuție 2 (d-f).

CAPITOLUL 7

Analiza Propagării Opiniilor la Nivel de Postare în Firele de Discuție din Forumuri

7.1. Definirea Problemei

Scopul nostru în analiza propagării opiniilor la nivel de postare este de a vizualiza, într-un mod simplificat, evoluția opiniilor exprimate într-un fir de discuție dintr-un forum pe măsură ce noi postări sunt adăugate. În acest scop, vom simplifica graful atașat firului de discuție, ținând cont de postările care conțin opinii. În graful final, fiecare nod rezultat din agregarea nodurilor vecine care au aceeași polaritate conține cuvintele de opinie care dau polaritatea respectivă.

Trebuie precizat că un cuvânt de opinie care exprimă un anumit sentiment (pozitiv, negativ sau neutru) poate avea o tărie a sentimentului diferită. Pentru a înțelege această observație vom da următorul exemplu: tăria sentimentului opiniei exprimate în propoziția „Camera este foarte luminată.” este diferită de tăria sentimentului opiniei exprimate în propoziția „Camera este luminată.”.

7.2. Arhitectura Abordării Propuse

Am împărțit problema propagării opiniilor la nivel de postare într-un fir de discuție dintr-un forum în următoarele etape, așa cum sunt prezentate mai jos și în figura 7.1:

- Preprocesarea fiecărei postări din graful asociat firului de discuție la momentul de timp t_τ , $\tau \in \mathbb{N}^*$. Vom nota acest graf cu $G_{DT}^0(t_\tau)(V_{DT}^0(t_\tau), E_{DT}^0(t_\tau))$;
- Filtrarea grafului $G_{DT}^0(t_\tau)(V_{DT}^0(t_\tau), E_{DT}^0(t_\tau))$ cu scopul de a elimina postările care conțin doar fapte și nu conțin opinii cu privire la subiectul firului de discuție. Notăm graful obținut la sfârșitul acestei etape cu $G_{DT}^1(t_\tau)(V_{DT}^1(t_\tau), E_{DT}^1(t_\tau))$;
- Calcularea sentimentului la nivel de fiecare postare din nodurile grafului $G_{DT}^1(t_\tau)(V_{DT}^1(t_\tau), E_{DT}^1(t_\tau))$ filtrat în etapa anterioară;
- Filtrarea grafului $G_{DT}^1(t_\tau)(V_{DT}^1(t_\tau), E_{DT}^1(t_\tau))$ cu scopul de a elimina postările cu sentiment neutru. Notăm graful obținut la sfârșitul acestei etape cu $G_{DT}^2(t_\tau)(V_{DT}^2(t_\tau), E_{DT}^2(t_\tau))$;
- Agregarea de noduri părinte-copil din graful $G_{DT}^2(t_\tau)(V_{DT}^2(t_\tau), E_{DT}^2(t_\tau))$ filtrat în etapa anterioară. Notăm graful obținut la sfârșitul acestei etape cu $G_{DT}^3(t_\tau)(V_{DT}^3(t_\tau), E_{DT}^3(t_\tau))$;
- Agregarea de noduri frați din graful $G_{DT}^3(t_\tau)(V_{DT}^3(t_\tau), E_{DT}^3(t_\tau))$. Notăm graful obținut la sfârșitul acestei etape cu $G_{DT}^4(t_\tau)(V_{DT}^4(t_\tau), E_{DT}^4(t_\tau))$.

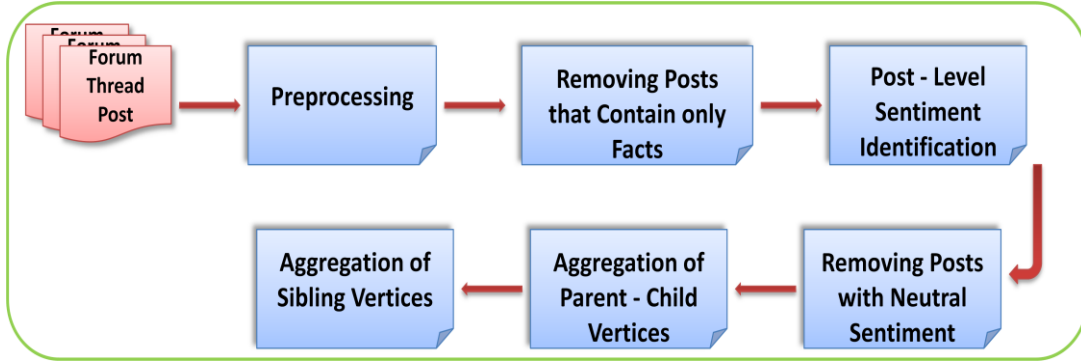


Figura 7.1: Arhitectura generală a problemei propagării opiniilor la nivel de postare într-un fir de discuție dintr-un forum.

Sistemul propus primește ca date de intrare firul de discuție dintr-un forum. În etapa de preprocesare, vom aplica sarcini specifice prelucrării limbajului natural (cum ar fi tokenizarea, etichetarea cu părți de vorbire și parsarea de dependențe) fiecărei postări din graful $G^0_{DT}(t_\tau)$ ($V^0_{DT}(t_\tau)$, $E^0_{DT}(t_\tau)$). Fiecare dintre aceste sarcini de prelucrare a limbajului natural au fost explicate în Capitolul 5.

În următoarele subsecțiuni, vom descrie fiecare etapă a arhitecturii prezentate în Figura 7.1. Aplicarea succesivă a acestor etape grafului inițial $G^0_{DT}(t_\tau)$ ($V^0_{DT}(t_\tau)$, $E^0_{DT}(t_\tau)$) va genera o serie de grafuri intermediare (a se vedea Algoritm 7.1). Relația dintre aceste grafuri poate fi descrisă astfel: $G^0_{DT}(t_\tau) \subset G^1_{DT}(t_\tau) \subset G^2_{DT}(t_\tau)$, $G^3_{DT}(t_\tau) \subset G^4_{DT}(t_\tau)$. Graful final $G^4_{DT}(t_\tau)$ ($V^4_{DT}(t_\tau)$, $E^4_{DT}(t_\tau)$) este folosit pentru a vizualiza evoluția opiniilor în mod simplificat pentru graful inițial $G^0_{DT}(t_\tau)$ ($V^0_{DT}(t_\tau)$, $E^0_{DT}(t_\tau)$), la momentul de timp t_τ , $\tau \in \mathbb{N}^*$.

Algoritm 7.1 (A7.1): Simplificarea Grafului Asociat Firului de Discuție la Momentul de Timp t_τ , $\tau \in \mathbb{N}^*$

Intrare: $G^0_{DT}(t_\tau)$ ($V^0_{DT}(t_\tau)$, $E^0_{DT}(t_\tau)$)

Ieșire: $G^4_{DT}(t_\tau)$ ($V^4_{DT}(t_\tau)$, $E^4_{DT}(t_\tau)$)

- 1: $G^1_{DT}(t_\tau)$ ($V^1_{DT}(t_\tau)$, $E^1_{DT}(t_\tau)$) $\leftarrow$ **Algoritm 7.2** ($G^0_{DT}(t_\tau)$ ($V^0_{DT}(t_\tau)$, $E^0_{DT}(t_\tau)$))
- 2: $G^2_{DT}(t_\tau)$ ($V^2_{DT}(t_\tau)$, $E^2_{DT}(t_\tau)$) $\leftarrow$ **Algoritm 7.3** ($G^1_{DT}(t_\tau)$ ($V^1_{DT}(t_\tau)$, $E^1_{DT}(t_\tau)$))
- 3: $G^3_{DT}(t_\tau)$ ($V^3_{DT}(t_\tau)$, $E^3_{DT}(t_\tau)$) $\leftarrow$ **Algoritm 7.4** ($G^2_{DT}(t_\tau)$ ($V^2_{DT}(t_\tau)$, $E^2_{DT}(t_\tau)$))
- 4: $G^4_{DT}(t_\tau)$ ($V^4_{DT}(t_\tau)$, $E^4_{DT}(t_\tau)$) $\leftarrow$ **Algoritm 7.5** ($G^3_{DT}(t_\tau)$ ($V^3_{DT}(t_\tau)$, $E^3_{DT}(t_\tau)$))

7.2.1. Eliminarea Postărilor care Conțin numai Fapte

Următoarea etapă în metoda propagării opiniilor la nivel de postare este eliminarea postărilor care conțin numai fapte din graful $G^0_{DT}(t_\tau)$ ($V^0_{DT}(t_\tau)$, $E^0_{DT}(t_\tau)$). În această etapă, vom elimina nodurile care nu sunt relevante în raport cu subiectul firului de discuție. Altfel spus, nodurile eliminate nu conțin opinii cu privire la subiectul firului de discuție. Celelalte noduri din graful $G^0_{DT}(t_\tau)$ ($V^0_{DT}(t_\tau)$, $E^0_{DT}(t_\tau)$) nu vor fi modificate.

Schița algoritmului pentru această etapă este arătată în Algoritm 7.2. Inițializăm graful de ieșire $G^1_{DT}(t_\tau)$ ($V^1_{DT}(t_\tau)$, $E^1_{DT}(t_\tau)$) pentru această etapă utilizând graful de intrare $G^0_{DT}(t_\tau)$ ($V^0_{DT}(t_\tau)$, $E^0_{DT}(t_\tau)$) (A7.2: 1-2). Apoi, vom aplica algoritmul de căutare în lățime pornind din nodul rădăcină v_1 , iar rezultatul obținut îl salvăm într-o listă (A7.2: 3). Pentru fiecare nod curent din listă, diferit de

nodul rădăcină v_1 , vom urma pașii următori. În primul rând, vom obține perechile (termen_substantival, cuvânt_opinie) pentru nodul curent, unde termen_substantival este legat semantic de unul dintre termenii substantivali care apar în denumirea subiectului firului de discuție (A7.2: 8). În cazul în care nu există perechi (termen_substantival, cuvânt_opinie) în nodul curent, vom elimina acest nod (A7.2: 9-17). Astfel, obținem nodul părinte al nodului curent (A7.2: 10). Pentru fiecare nod copil al nodului curent, vom crea o legătură între acesta și nodul părinte al nodului curent (A7.2: 14). În cele din urmă, vom elimina nodul curent din mulțimea $V_{DT}^1(t_r)$ (A7.2: 15).

Algoritm 7.2 (A7.2): Eliminarea Postărilor care Conțin numai Fapte

Intrare: $G_{DT}^0(t_r)(V_{DT}^0(t_r), E_{DT}^0(t_r))$
Ieșire: $G_{DT}^1(t_r)(V_{DT}^1(t_r), E_{DT}^1(t_r))$

```

1:  $V_{DT}^1(t_r) \leftarrow V_{DT}^0(t_r)$ 
2:  $E_{DT}^1(t_r) \leftarrow E_{DT}^0(t_r)$ 
3:  $\mathcal{M} \leftarrow \text{BreadthFirstSearch}(v_1)$ 
4: for each node crtNode in  $\mathcal{M}$  do
5:   if crtNode =  $v_1$  then
6:     continue
7:   endif
8:    $\Omega \leftarrow \text{NounTermVocabularyConstruction}(\text{crtNode})$ 
9:   if  $\Omega = \emptyset$  then
10:    parentNode  $\leftarrow \text{GetParentNode}(\text{crtNode})$ 
11:     $\mathcal{N} \leftarrow \text{GetChildrenNodes}(\text{crtNode})$ 
12:    for each node childNode in  $\mathcal{N}$  do
13:       $E_{DT}^1(t_r) = E_{DT}^1(t_r) \setminus (\text{childNode}, \text{crtNode})$ 
14:       $E_{DT}^1(t_r) = E_{DT}^1(t_r) \cup (\text{childNode}, \text{parentNode})$ 
15:       $V_{DT}^1(t_r) = V_{DT}^1(t_r) \setminus \{\text{crtNode}\}$ 
16:    end for
17:  endif
18: endfor
    
```

7.2.2. Identificarea Sentimentului la Nivel de Postare

Pentru a determina sentimentul exprimat de o postare, noi ținem seama de tăria sentimentelor exprimate de cuvintele de opinie din respectiva postare. Fie $p \in P_{DT}(t_r)$ o postare într-un fir de discuție. Scorul asociat sentimentului postării p este dat de următoarea formulă:

$$\begin{aligned}
 & \text{postSentimentScore}(p) \\
 &= \frac{\sum_{i=1}^4 \wedge w \in S_J^i(p) \wedge w \in S_R^i(p) \text{ score}(w) * \lambda_i + \sum_{w \in S_V(p)} \text{score}(w) * \lambda_4}{\sum_{i=1}^4 |S_J^i(p)| + \sum_{i=1}^4 |S_R^i(p)| + |S_V(p)|}
 \end{aligned} \tag{7.1}$$

unde:

- $\text{score}(w)$ reprezintă scorul returnat pentru cuvântul de opinie w de un vocabular de opinii;
- $S_J^1(p)$ reprezintă mulțimea de adjective la gradul superlativ din postarea p ;
- $S_J^2(p)$ reprezintă mulțimea de adjective la gradul comparativ de superioritate din postarea p ;

- $S^3_J(p)$ reprezintă mulțimea de adjective la gradul comparativ de inferioritate din postarea p ;
- $S^4_J(p)$ reprezintă mulțimea de adjective la celelalte grade de comparație din postarea p ;
- $S^1_R(p)$ reprezintă mulțimea de adverbe la gradul superlativ din postarea p ;
- $S^2_R(p)$ reprezintă mulțimea de adverbe la gradul comparativ de superioritate din postarea p ;
- $S^3_R(p)$ reprezintă mulțimea de adverbe la gradul comparativ de inferioritate din postarea p ;
- $S^4_R(p)$ reprezintă mulțimea de adverbe la celelalte grade de comparație din postarea p ;
- $S_V(p)$ reprezintă mulțimea de verbe din postarea p ;
- $|S|$ reprezintă cardinalul mulțimii S .

Pentru variabilele λ_1 , λ_2 , λ_3 și λ_4 , am setat valorile 0.9, 0.6, -0.6, respectiv 0.3. Postarea $p \in P_{DT}(t_r)$ exprimă un sentiment pozitiv / negativ în cazul în care $postScore(p) \in (0, +1]$ / $postScore(p) \in [-1, 0)$.

7.2.3. Eliminarea Postărilor cu Sentiment Neutru

Următoarea etapă în metoda propagării opiniilor la nivel de postare este eliminarea postărilor cu sentiment neutru din graful $G^1_{DT}(t_r)(V^1_{DT}(t_r), E^1_{DT}(t_r))$. Cu alte cuvinte, nodurile eliminate nu conțin opinii cu sentimente pozitive sau negative cu privire la subiectul firului de discuție, ci doar opinii cu sentiment neutru sau fapte.

Schița algoritmului pentru această etapă este arătată în Algoritmul 7.3. Mai întâi, inițializăm graful de ieșire $G^2_{DT}(t_r)(V^2_{DT}(t_r), E^2_{DT}(t_r))$ utilizând graful de intrare $G^1_{DT}(t_r)(V^1_{DT}(t_r), E^1_{DT}(t_r))$ (A7.3: 1-2). Apoi, vom aplica algoritmul de căutare în lățime pornind din nodul rădăcină v_1 și salvăm rezultatul într-o listă (A7.3: 3). Pentru fiecare nod curent din listă, diferit de nodul rădăcină v_1 , vom urma pașii următori. În primul rând, calculăm scorul de sentiment pentru nodul curent utilizând formula prezentată în secțiunea 7.2.2. (A7.3: 8). Dacă acest scor de sentiment nu are valoarea zero, atunci obținem nodul părinte al nodului curent (A7.3: 10) și vom crea o legătură între fiecare nod copil al nodului curent și nodul părinte al nodului curent (A7.3: 14). În cele din urmă, vom elimina nodul curent din mulțimea $V^2_{DT}(t_r)$ (A7.3: 15).

Algoritm 7.3 (A7.3): Eliminarea Postărilor cu Sentiment Neutru

Intrare: $G^1_{DT}(t_r)(V^1_{DT}(t_r), E^1_{DT}(t_r))$

Ieșire: $G^2_{DT}(t_r)(V^2_{DT}(t_r), E^2_{DT}(t_r))$

- 1: $V^2_{DT}(t_r) \leftarrow V^1_{DT}(t_r)$
- 2: $E^2_{DT}(t_r) \leftarrow E^1_{DT}(t_r)$
- 3: $\mathcal{M} \leftarrow \text{BreadthFirstSearch}(v_1)$
- 4: **for each** node crtNode in $\mathcal{M}$ **do**
- 5: **if** crtNode = v_1 **then**
- 6: **continue**
- 7: **endif**
- 8: sentimentScore $\leftarrow \text{GetNodeSentiment}(\text{crtNode})$
- 9: **if** sentimentScore = 0 **then**
- 10: parentNode $\leftarrow \text{GetParentNode}(\text{crtNode})$
- 11: $\mathcal{N} \leftarrow \text{GetChildrenNodes}(\text{crtNode})$
- 12: **for each** node childNode in $\mathcal{N}$ **do**

```

13:       $E_{DT}^1(t_\tau) = E_{DT}^2(t_\tau) \setminus (\text{childNode}, \text{crtNode})$ 
14:       $E_{DT}^2(t_\tau) = E_{DT}^2(t_\tau) \cup (\text{childNode}, \text{parentNode})$ 
15:       $V_{DT}^2(t_\tau) = V_{DT}^2(t_\tau) \setminus \{\text{crtNode}\}$ 
16:  end for
17: endif
18: endfor
    
```

7.2.4. Agregarea de Noduri Părinte-Copil

Următoarea etapă în metoda propagării opiniilor la nivel de postare este agregarea de noduri părinte-copil. Această etapă de agregare primește ca date de intrare graful $G_{DT}^2(t_\tau)(V_{DT}^2(t_\tau), E_{DT}^2(t_\tau))$, unde $G_{DT}^2(t_\tau)$ este ieșirea etapei anterioare. Agregarea de noduri părinte-copil are loc în conformitate cu următoarea definiție:

Definiția 7.1 (Agregarea de noduri părinte-copil). *Fiind date la timpul $t_\tau, \tau \in \mathbb{N}^*$, firul de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum și graful $G_{DT}^2(t_\tau)(V_{DT}^2(t_\tau), E_{DT}^2(t_\tau))$ asociat firului de discuție, două noduri $v_i, v_k \in V_{DT}^2(t_\tau)$, $v_i = (v_i^p, v_i^u, v_i^{tm}, v_i^{op})$, $v_k = (v_k^p, v_k^u, v_k^{tm}, v_k^{op})$, vor fi colapsate dacă $(v_i, v_k) \in E_{DT}^2(t_\tau)$ și nodurile v_i, v_k au aceeași polaritate (pozitivă sau negativă). Rezultatul agregării nodurilor $v_i, v_k \in V_{DT}^2(t_\tau)$ este un singur nod $v_r = (v_r^p, v_r^u, v_r^{tm}, v_r^{op}) \in V_{DT}^3(t_\tau)$ caracterizat prin: $v_r^p = v_i^p \cup v_k^p$, $v_r^u = v_i^u \cup v_k^u$, $v_r^{tm} = v_i^{tm} \cup v_k^{tm}$ și $v_r^{op} = v_i^{op} \cup v_k^{op}$.*

În continuare, vom da un exemplu pentru a ilustra această definiție. În Figura 7.2(a), nodul v_l este un răspuns la nodul v_k , nodul v_k este un răspuns la nodul v_j , iar nodul v_j este un răspuns la nodul v_i . Nodurile v_i, v_j și v_k au aceeași polaritate pozitivă și vor fi agregate în conformitate cu definiția 7.1. Rezultatul agregării este nodul v'_{ijk} de polaritate pozitivă, iar nodul v_l este un răspuns la nodul v'_{ijk} . În schimb, în Figura 7.2 (b), pe calea de la nodul v_l la nodul v_i există o alternanță de noduri de polarități pozitive și negative, iar definiția 7.1 nu poate fi aplicată în acest caz.

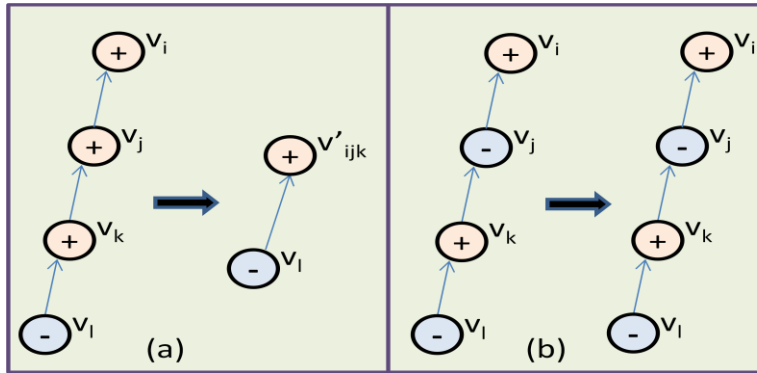


Figura 7.2: (a) Exemplu posibil de agregare de noduri părinte-copil; (b) Exemplu când agregarea de noduri părinte-copil nu este posibilă.

Schița algoritmului care realizează agregarea de noduri părinte-copil și transformă graful $G_{DT}^2(t_\tau)(V_{DT}^2(t_\tau), E_{DT}^2(t_\tau))$ în graful $G_{DT}^3(t_\tau)(V_{DT}^3(t_\tau), E_{DT}^3(t_\tau))$ este dată în algoritmul 7.4.

Algoritm 7.4 (A7.4): Agregarea de Noduri Părinte-Copil

Intrare: $G_{DT}^2(t_\tau)(V_{DT}^2(t_\tau), E_{DT}^2(t_\tau))$

Ieșire: $G^3_{DT}(t_\tau)(V^3_{DT}(t_\tau), E^3_{DT}(t_\tau))$

```

1:  $V^3_{DT}(t_\tau) \leftarrow V^2_{DT}(t_\tau)$ 
2:  $E^3_{DT}(t_\tau) \leftarrow E^2_{DT}(t_\tau)$ 
3:  $\mathcal{M} \leftarrow \text{BreadthFirstSearch}(v_1)$ 
4: for each node crtNode in  $\mathcal{M}$  do
5:   parentNode  $\leftarrow \text{GetParentNode}(\text{crtNode})$ 
6:   if crtNode =  $v_1$  or parentNode =  $v_1$  then
7:     continue
8:   end if
9:   crtNodeSentimentScore  $\leftarrow \text{GetNodeSentiment}(\text{crtNode})$ 
10:  parentNodeSentimentScore  $\leftarrow \text{GetNodeSentiment}(\text{parentNode})$ 
11:  if crtNodeSentimentScore * parentNodeSentimentScore > 0 then
12:     $\mathcal{N} \leftarrow \text{GetChildrenNodes}(\text{crtNode})$ 
13:    for each node childNode in  $\mathcal{N}$  do
14:       $E^3_{DT}(t_\tau) = E^3_{DT}(t_\tau) \setminus (\text{childNode}, \text{crtNode})$ 
15:       $E^3_{DT}(t_\tau) = E^3_{DT}(t_\tau) \cup (\text{childNode}, \text{parentNode})$ 
16:      InformationUpdate(parentNode, crtNode)
17:       $V^3_{DT}(t_\tau) = V^3_{DT}(t_\tau) \setminus \{\text{crtNode}\}$ 
18:    end for
19:  end if
20: end for
```

Pentru graful $G^3_{DT}(t_\tau)(V^3_{DT}(t_\tau), E^3_{DT}(t_\tau))$, putem scrie următoarea definiție:

Definiția 7.2 Fiind date la timpul t_τ , $\tau \in \mathbb{N}^*$, firul de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum, precum și graful $G^2_{DT}(t_\tau)(V^2_{DT}(t_\tau), E^2_{DT}(t_\tau))$ asociat firului de discuție și obținut conform algoritmului 7.3, atunci asociem firului de discuție graful orientat $G^3_{DT}(t_\tau)(V^3_{DT}(t_\tau), E^3_{DT}(t_\tau))$ obținut după aplicarea algoritmului 7.4, unde:

- $V^3_{DT}(t_\tau) = \{v'_j \mid v'_j = (\cup v^p_{j,l}, \cup v^u_{j,l}, \cup v^{tm}_{j,l}, \cup v^{op}_{j,l}), v^p_{j,l} \subset P_{DT}(t_\tau), v^u_{j,l} \subset U_{DT}(t_\tau), v^{tm}_{j,l} \in \mathbb{N}, v^{op}_{j,l} \subset OS^d_{DT}\}$ reprezintă mulțimea de noduri în graful $G^3_{DT}(t_\tau)$. Un nod v'_j în $V^3_{DT}(t_\tau)$ conține o mulțime de postări $\cup v^p_{j,l}$ scrise de o mulțime de utilizatori $\cup v^u_{j,l}$ la momentele de timp $\cup v^{tm}_{j,l}$;
- $E^3_{DT}(t_\tau) = \{e'_1, e'_2, \dots, e'_r\}$ reprezintă mulțimea de arce în graful $G^3_{DT}(t_\tau)$ astfel încât, dacă $e' = (v'_i, v'_j) \in E^3_{DT}(t_\tau)$, mulțimea de postări corespunzătoare nodului v'_i este un răspuns la mulțimea de postări corespunzătoare nodului v'_j , iar nodurile v'_i, v'_j au polarități opuse (nodul v'_i exprimă un sentiment pozitiv și nodul v'_j exprimă un sentiment negativ sau nodul v'_i exprimă un sentiment negativ și nodul v'_j exprimă un sentiment pozitiv).

7.2.5. Agregarea de Noduri Frați

Următoarea etapă în metoda propagării opiniilor la nivel de postare este agregarea de noduri care au nodul părinte comun. Această etapă de agregare primește ca date de intrare graful $G^3_{DT}(t_\tau)(V^3_{DT}(t_\tau), E^3_{DT}(t_\tau))$, unde $G^3_{DT}(t_\tau)$ este ieșirea etapei anterioare. Agregarea de noduri frați are loc în conformitate cu următoarea definiție:

Definiția 7.3 (Agregarea de noduri frați). Fiind date la timpul t_τ , $\tau \in \mathbb{N}^*$, firul de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum și graful $G^3_{DT}(t_\tau)(V^3_{DT}(t_\tau), E^3_{DT}(t_\tau))$ asociat firului de discuție, atunci două noduri $v_i, v_k \in V^3_{DT}(t_\tau)$, $v_i = (v_i^p, v_i^u, v_i^{tm}, v_i^{op})$, $v_k = (v_k^p, v_k^u, v_k^{tm}, v_k^{op})$, vor fi colapsate dacă există nodul $v_j \in V^3_{DT}(t_\tau)$ astfel încât $(v_i, v_j) \in E^3_{DT}(t_\tau)$, $(v_k, v_j) \in E^3_{DT}(t_\tau)$ și nodurile v_i, v_k au aceeași polaritate (pozitivă sau negativă). Rezultatul agregării nodurilor v_i, v_k este un singur nod $v_r = (v_r^p, v_r^u, v_r^{tm}, v_r^{op}) \in V^4_{DT}(t_\tau)$ caracterizat prin: $v_r^p = v_i^p \cup v_k^p$, $v_r^u = v_i^u \cup v_k^u$, $v_r^{tm} = v_i^{tm} \cup v_k^{tm}$ și $v_r^{op} = v_i^{op} \cup v_k^{op}$.

În continuare, vom da un exemplu pentru a ilustra această definiție. În Figura 7.3(c), nodul v_l este un răspuns la nodul v_k , iar nodul v_k este un răspuns la nodul v_i . Ambele noduri v_i și v_j au aceeași polaritate pozitivă și același nod părinte v_0 . Aplicând definiția 7.3 nodurilor v_i și v_j , obținem nodul v'_{ij} de polaritate pozitivă, unde nodul v'_{ij} este un răspuns la nodul v_0 . În schimb, în Figura 7.3(d), cele două noduri v_i și v_j au polaritate opusă, iar definiția 7.3 nu poate fi aplicată în acest caz.

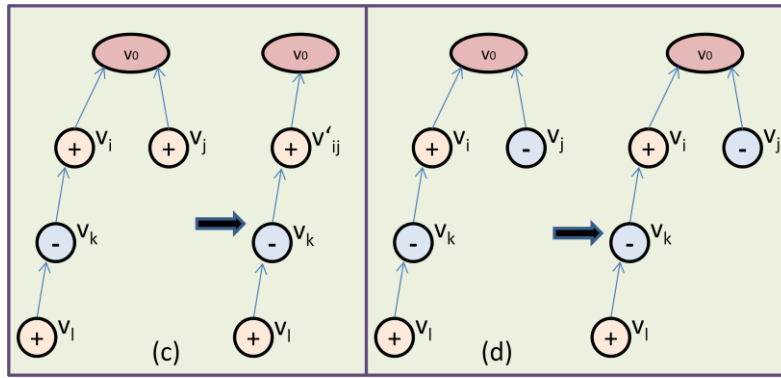


Figura 7.3: (c) Exemplu posibil de agregare de noduri frați; (d) Exemplu când agregarea de noduri frați nu este posibilă.

Schița algoritmului care realizează agregarea de noduri frați și transformă graful $G^3_{DT}(t_\tau)(V^3_{DT}(t_\tau), E^3_{DT}(t_\tau))$ în graful $G^4_{DT}(t_\tau)(V^4_{DT}(t_\tau), E^4_{DT}(t_\tau))$ este dată în algoritmul 7.5.

Algoritm 7.5 (A7.5): Agregarea de Noduri Frați

Intrare: $G^3_{DT}(t_\tau)(V^3_{DT}(t_\tau), E^3_{DT}(t_\tau))$

Ieșire: $G^4_{DT}(t_\tau)(V^4_{DT}(t_\tau), E^4_{DT}(t_\tau))$

- 1: $V^4_{DT}(t_\tau) \leftarrow V^3_{DT}(t_\tau)$
- 2: $E^4_{DT}(t_\tau) \leftarrow E^3_{DT}(t_\tau)$
- 3: $\mathcal{M} \leftarrow \{v_1\}$
- 4: **while** $\mathcal{M} \neq \emptyset$
- 5: positiveNodesList $\leftarrow \emptyset$
- 6: negativeNodesList $\leftarrow \emptyset$
- 7: crtNode $\leftarrow \text{RemoveNode}(\mathcal{M})$
- 8: $\mathcal{N} \leftarrow \text{GetChildrenNodes}(\text{crtNode})$
- 9: **for each** node childNode in $\mathcal{N}$ **do**
- 10: childSentimentScore $\leftarrow \text{GetNodeSentiment}(\text{childNode})$
- 11: **if** childSentimentScore > 0 **then**
- 12: **if** positiveNodesList = $\emptyset$ **then**
- 13: positiveNode = childNode

```

14:      else
15:          positiveNode = positiveNode ∪ { childNode }
16:           $V^4_{DT}(t_\tau) = V^4_{DT}(t_\tau) \setminus \{crtNode\}$ 
17:      end if
18:  end if
19:  if childSentimentScore < 0 then
20:      if negativeNode = ∅ then
21:          negativeNode = childNode
22:      else
23:          negativeNode = negativeNode ∪ { childNode }
24:           $V^4_{DT}(t_\tau) = V^4_{DT}(t_\tau) \setminus \{crtNode\}$ 
25:      end if
26:  end if
27: end for
28: if positiveNodesList != ∅ then
29:     for each node childNode in positiveNodesList do
30:         AddNode( $\mathcal{M}$ , childNode);
31:     end for
32: end if
33: if negativeNodesList != ∅ then
34:     for each node childNode in negativeNodesList do
35:         AddNode( $\mathcal{M}$ , childNode);
36:     end for
37: end if
38: end while

```

Pentru graful $G^4_{DT}(t_\tau)(V^4_{DT}(t_\tau), E^4_{DT}(t_\tau))$, putem scrie următoarea definiție:

Definiția 7.4 Fiind date la timpul t_τ , $\tau \in \mathbb{N}^*$, firul de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum și graful $G^3_{DT}(t_\tau)(V^3_{DT}(t_\tau), E^3_{DT}(t_\tau))$ asociat firului de discuție obținut conform algoritmului 7.4, atunci asociem firului de discuție graful orientat $G^4_{DT}(t_\tau)(V^4_{DT}(t_\tau), E^4_{DT}(t_\tau))$ obținut după aplicarea algoritmului 7.5, unde:

- $V^4_{DT}(t_\tau) = \{v'_j \mid v'_j = (\cup v^p_{j,l}, \cup v^u_{j,l}, \cup v^{tm}_{j,l}, \cup v^{op}_{j,l}), v^p_{j,l} \subset P_{DT}(t_\tau), v^u_{j,l} \subset U_{DT}(t_\tau), v^{tm}_{j,l} \in \mathbb{N}, v^{op}_{j,l} \subset OS^d_{DT}\}$ reprezintă mulțimea de noduri în graful $G^4_{DT}(t_\tau)$. Un nod v'_j în $V^4_{DT}(t_\tau)$ conține o mulțime de postări $\cup v^p_{j,l}$ scrise de o mulțime de utilizatori $\cup v^u_{j,l}$ la momentele de timp $\cup v^{tm}_{j,l}$.
- $E^4_{DT}(t_\tau) = \{e'_1, e'_2, \dots, e'_s\}$ reprezintă mulțimea de arce în graful $G^4_{DT}(t_\tau)$ astfel încât, dacă $e' = (v'_i, v'_j) \in E^4_{DT}(t_\tau)$, mulțimea de postări corespunzătoare nodului v'_i este un răspuns la mulțimea de postări corespunzătoare nodului v'_j și nodurile v'_i, v'_j au polarități opuse. Mai mult, dacă $v'_i, v'_j, v'_k \in V^4_{DT}(t_\tau)$, astfel încât $(v'_i, v'_k) \in E^4_{DT}(t_\tau)$, $(v'_j, v'_k) \in E^4_{DT}(t_\tau)$, atunci nodurile v'_i și v'_j au polarități opuse (nodul v'_i exprimă un sentiment pozitiv și nodul v'_j exprimă un sentiment negativ sau nodul v'_i exprimă un sentiment negativ și nodul v'_j exprimă un sentiment pozitiv).

CAPITOLUL 8

Analiza Influenței Opiniilor în Firele de Discuție din Forumuri

8.1. Introducere

Utilizatorii care interacționează pe forumurile de discuții au diverse motivații: să converseze, să dobândească cunoștințe, să promoveze diverse produse etc. Un alt aspect important al interacțiunilor online este acela că oamenii își pot spune opiniile pe diferite teme. Aceeași opinie despre o anumită temă poate fi împărtășită de unul sau mai mulți participanți. Cu toate acestea, există și participanți care au opinii contrare.

De regulă, în discuțiile online apare fenomenul de influență între opiniile participanților. Pe de o parte, un participant care și-a format deja o opinie și-o poate modifica ca urmare a influenței venite din partea altui participant / altor participanți. Pe de altă parte, fenomenul de influență ar putea avea ca efect formarea de noi opinii în cazul participanților care nu aveau anterior nicio opinie formată.

Analiza problemei propagării de opinii în mediile sociale este complexă și trebuie să țină cont de influența dintre participanți. Influența socială este definită ca puterea exercitată de către un individ asupra unui individ țintă cu scopul de a-i schimba gândurile, opiniile, comportamentul. Influența socială este rezultatul interacțiunilor dintre utilizatorii unui mediu social. Există două tipuri de influență într-o rețea de socializare online: influența directă, ce are loc între două persoane care sunt conectate în mod direct [39] și influența indirectă, manifestată între două persoane care nu sunt conectate direct [40].

În acest capitol, propunem o analiză a fenomenului de influență în firele de discuție din forumuri, luând în considerare opiniile exprimate de participanții la discuție. Ipoteza care stă la baza cercetării noastre este că, în firele de discuție, opiniile participanților se pot modifica de-a lungul timpului ca urmare a interacțiunilor dintre ei sau cu noii participanți. Astfel, o postare trimisă de un utilizator pe un fir de discuție poate avea o influență asupra celorlalte postări trimise ca reacție de către alți utilizatori.

Pentru a rezolva această problemă, abordarea noastră se bazează pe sarcini specifice de prelucrare a limbajului natural și pe analiza semantică latentă [10, p. 243] ca măsură de similaritate între două postări. Rezultatele arată că, în aceeași cale de discuție, toate postările anterioare ce conțin opinii au o influență într-o proporție diferită asupra unei noi postări.

8.2. Formularea Problemei

Definiția 8.1 (Cale de Discuție). Fiind dat la timpul t_τ , $\tau \in \mathbb{N}^*$, graful $G_{DT}(t_\tau)(V_{DT}(t_\tau), E_{DT}(t_\tau))$ asociat firului de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum online, putem defini o cale de discuție $V_{DC}(v_{k1}, v_{kl})$ ca o mulțime de noduri $\{v_{k1}, v_{k2}, \dots, v_{kl}\}$ astfel încât $(v_{ki}, v_{k(i+1)}) \in E_{DT}(t_\tau)$, pentru $i \in \{1, 2, \dots, l-1\}$ și $v_{kl} = v_1$, unde nodul rădăcină $v_1 \in V_{DT}(t_\tau)$ corespunde primei postări trimise de către utilizatorul $u_1 \in U_{DT}(t_\tau)$ pe acel fir de discuție, la momentul de timp t_1 .

Fie $C(u, t_\tau) = \{p \in P_{DT}(t_\tau) | (u, p) \in R_{UP}(t_\tau)\}$ mulțimea postărilor scrise de către utilizatorul $u \in U_{DT}(t_\tau)$ până la momentul de timp t_τ , $\tau \in \mathbb{N}^*$. Să considerăm ordonarea mulțimii $C(u, t_\tau)$ după

momentele de timp când postările utilizatorului $u \in U_{DT}(t_\tau)$ au fost trimise pe firul de discuție ca fiind $p_u : \mathcal{N} \rightarrow C(u, t_\tau)$. Luând în considerare ordonarea mulțimii $C(u, t_\tau)$, rescriem această mulțime ca $\{p_u^k \in P_{DT}(t_\tau) \mid (u, p_u^k) \in R_{UP}(t_\tau), k \in \mathcal{N}\}$.

Fie $C'(u, t_\tau) = \{p_u^k \in P_{DT}(t_\tau) \mid (u, p_u^k) \in R_{UP}(t_\tau), k \in \mathcal{N}^* \text{ și } postSentiment(p_u^k) \neq postSentiment(p_u^0)\}$ mulțimea de postări trimise de către utilizatorul $u \in U_{DT}(t_\tau)$ la momentele de timp $t_{\tau i} > t_{\tau l}$ și având o polaritate opusă primei postări trimise de același utilizator u la momentul de timp $t_{\tau l}$. Procedura $postSentiment$ calculează orientarea semantică a fiecărei postări. În cele din urmă, ținând seama de toate explicațiile de mai sus, putem da următoarea definiție:

Definiția 8.2 (Problema Analizei Influenței Opiniilor în Firele de Discuție din Forumuri). *Fiind dat la timpul t_τ , $\tau \in \mathcal{N}^*$, graful $G_{DT}(t_\tau)(V_{DT}(t_\tau), E_{DT}(t_\tau))$ asociat firului de discuție $(T_{DT}, S_{DT}, U_{DT}(t_\tau), P_{DT}(t_\tau), R_{DT}(t_\tau))$ dintr-un forum, un utilizator $u \in U_{DT}(t_\tau)$ își schimbă opinia dacă condiția $C'(u, t_\tau) \neq \emptyset$ este adevărată. În caz afirmativ, utilizatorul $u \in U_{DT}(t_\tau)$ își schimbă opinia în postarea p_u^j dată de următoarea ecuație:*

$$(p_u^j, p_w^k) = \arg \max_{p_u^j \in C'(u, t_\tau), p_w^k \in V_{DC}(p_{u0}^0, p_u^j)} (similarity(p_u^j, p_w^k)) \quad (8.1)$$

unde $similarity(p_u^j, p_w^k)$ măsoară similaritatea semantică dintre două postări. Ecuația 8.1 indică faptul că utilizatorul $u \in U_{DT}(t_\tau)$ își schimbă opinia în postarea p_u^j deoarece a fost influențat de postarea p_w^k trimisă de utilizatorul w .

8.3. Arhitectura Abordării Propuse

Sistemul care implementează metoda propusă de noi este compus din câteva module, care sunt prezentate în Figura 8.1.

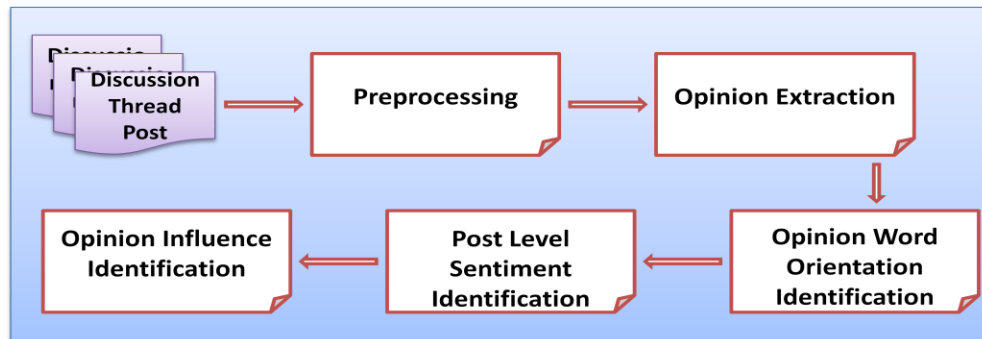


Figura 8.1: Arhitectura generală a analizei influenței opiniilor participanților într-un fir de discuție dintr-un forum.

8.4. Experimente

Experimentele din această secțiune sunt efectuate folosind un fir de forum¹ din corpusul creat de Walker ș.a. [38]. Acest fir de discuție conține 192 de postări scrise de 20 de utilizatori. În Figura 8.2, este reprezentată interfața de vizualizare a acestui fir de discuție sub forma unui graf. În acest

¹ <http://www.4forums.com/political/religion-debates/3742-religion-opium-people.html>

scop, am folosit biblioteca Graphstream¹. GraphStream este o bibliotecă în Java, dezvoltată pentru a modela grafurile dinamice. Nodul rădăcină din graf este reprezentat prin culoarea albastră. Celelalte noduri din graf sunt reprezentate prin culoarea gri. Mai mult decât atât, un identificator este atribuit fiecărui nod al grafului. Identificatorii pentru noduri sunt valori numerice mai mari sau egale cu 1. Valoarea 1 este asociată cu prima postare din firul de discuție, adică cu nodul rădăcină al grafului asociat firului de discuție.

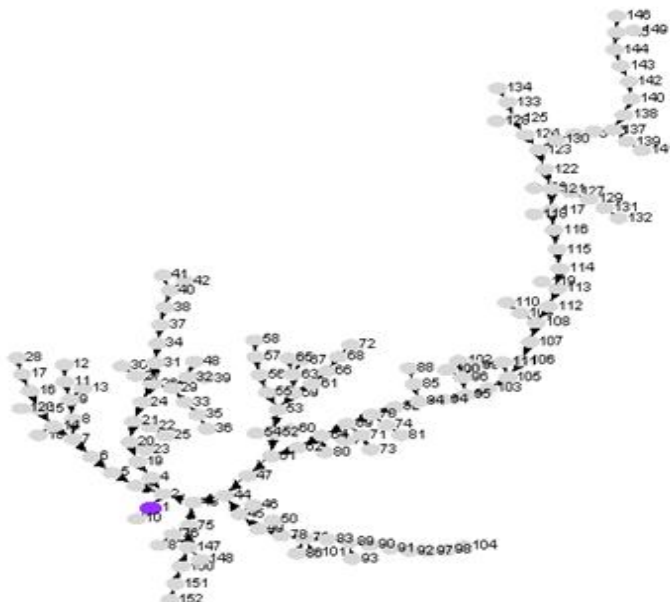


Figura 8.2. Vizualizarea unui fir de forum² sub forma unui graf.

În Figura 8.3, am reprezentat interfața aplicației care permite vizualizarea sentimentului exprimat de fiecare postare trimisă de un participant într-un fir de discuție. Această interfață are două zone: în stânga, o zonă care permite interacțiunea între un utilizator și interfața aplicației, iar în dreapta, o zonă care permite vizualizarea rezultatelor.

În partea de sus a zonei din stânga, se poate citi subiectul firului de discuție. Apoi, putem selecta termenii substantivali din toate postările $P_{DT}(t_r)$. Acești termeni substantivali sunt similari din punct de vedere semantic cu subiectul firului de discuție, iar utilizatorii din mulțimea $U_{DT}(t_r)$ și-au exprimat opiniile cu privire la acești termeni substantivali de-a lungul firului de discuție. În partea de jos a zonei din stânga, se poate selecta un utilizator din mulțimea $U_{DT}(t_r)$.

În zona din dreapta, am reprezentat toate postările $C(u, t_r)$ trimise de un participant selectat $u \in U_{DT}(t_r)$ până la momentul de timp t_r , $t_r \in \mathbb{N}^*$. Postările din mulțimea $C(u, t_r)$ sunt reprezentate în ordine cronologică. În plus, se poate observa sentimentul asociat cu fiecare postare din mulțimea $C(u, t_r)$, unde valoarea +1 indică un sentiment pozitiv, valoarea 0 indică un sentiment neutru, iar valoarea -1 indică un sentiment negativ.

În Figura 8.3, am reprezentat sentimentul asociat cu fiecare postare trimisă de către participantul 11, pentru fiecare dintre cele patru vocabulare de opinii utilizate. Se poate observa că toate postările din mulțimea $C(u, t_r) = \{p_{11}^{26}, p_{11}^{36}, p_{11}^{44}, p_{11}^{51}, p_{11}^{54}, p_{11}^{70}, p_{11}^{88}, p_{11}^{95}\}$ exprimă un sentiment opus față de sentimentul asociat postării p_{11}^{21} .

¹ <http://graphstream-project.org>

² <http://www.4forums.com/political/religion-debates/3742-religion-opium-people.html>

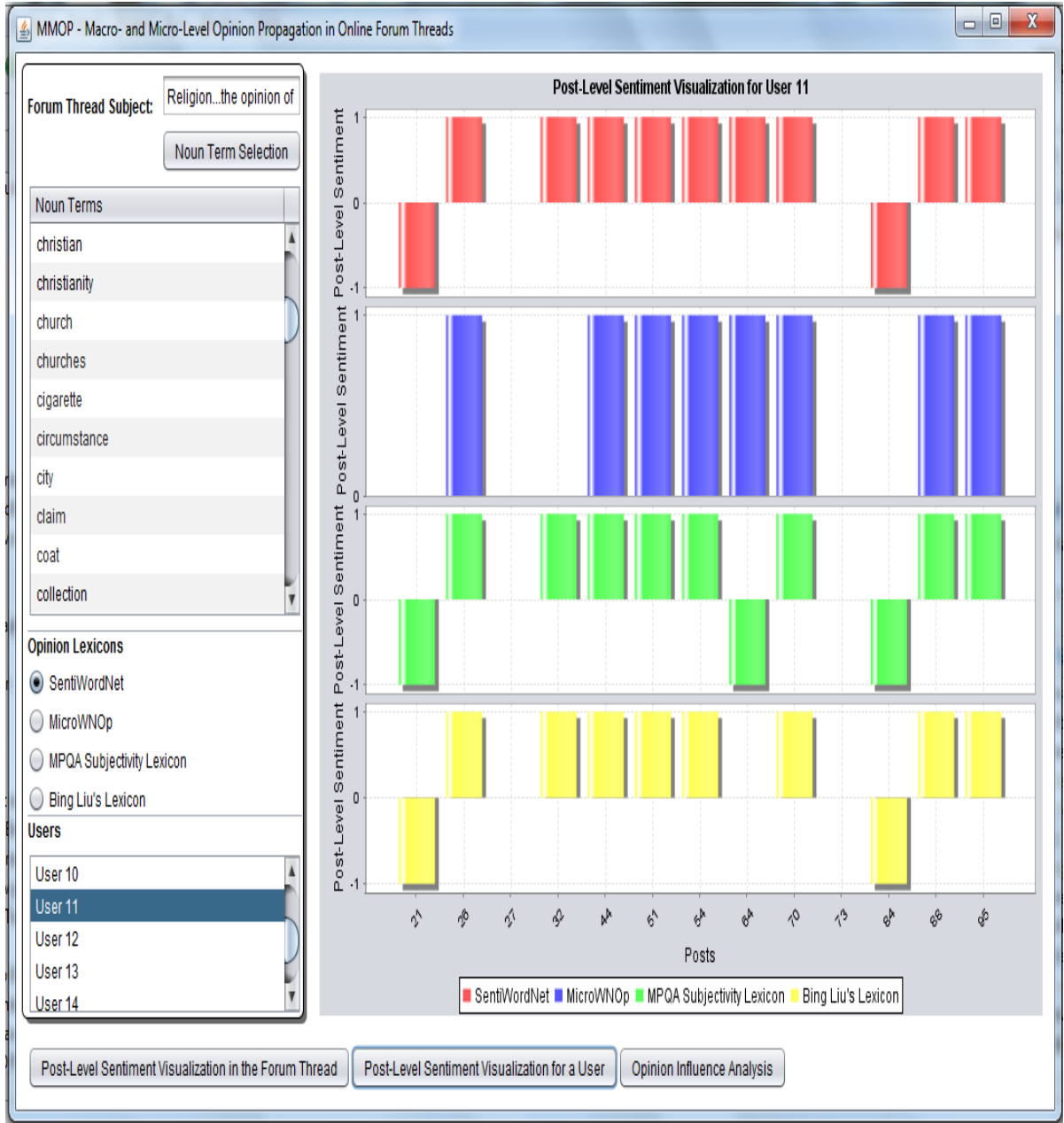


Figura 8.3. Interfața aplicației care permite vizualizarea sentimentului exprimat de fiecare postare trimisă de către Participantul 11 în firul de discuție considerat.

În Figura 8.4, am reprezentat interfața aplicației care permite analiza influenței opiniei pentru un participant u selectat din mulțimea $U_{DI}(t_r)$. În acest caz, interfața aplicației este împărțită în trei zone. Zona de sus arată lista $C(u, t_r)$ corespunzătoare participantului selectat u . În zona de jos, am reprezentat calea de discuție $V_{DC}(p_u^k, v_1)$, ce corespunde unei postări p_u^k selectate din mulțimea $C(u, t_r)$. În zona centrală, am reprezentat grafic valoarea influenței dintre prima postare trimisă de către utilizatorul $u \in U_{DI}(t_r)$ și fiecare postare din mulțimea $V_{DC}(p_u^k, v_1)$. Valoarea influenței între două postări este calculată folosind analiza semantică latentă.

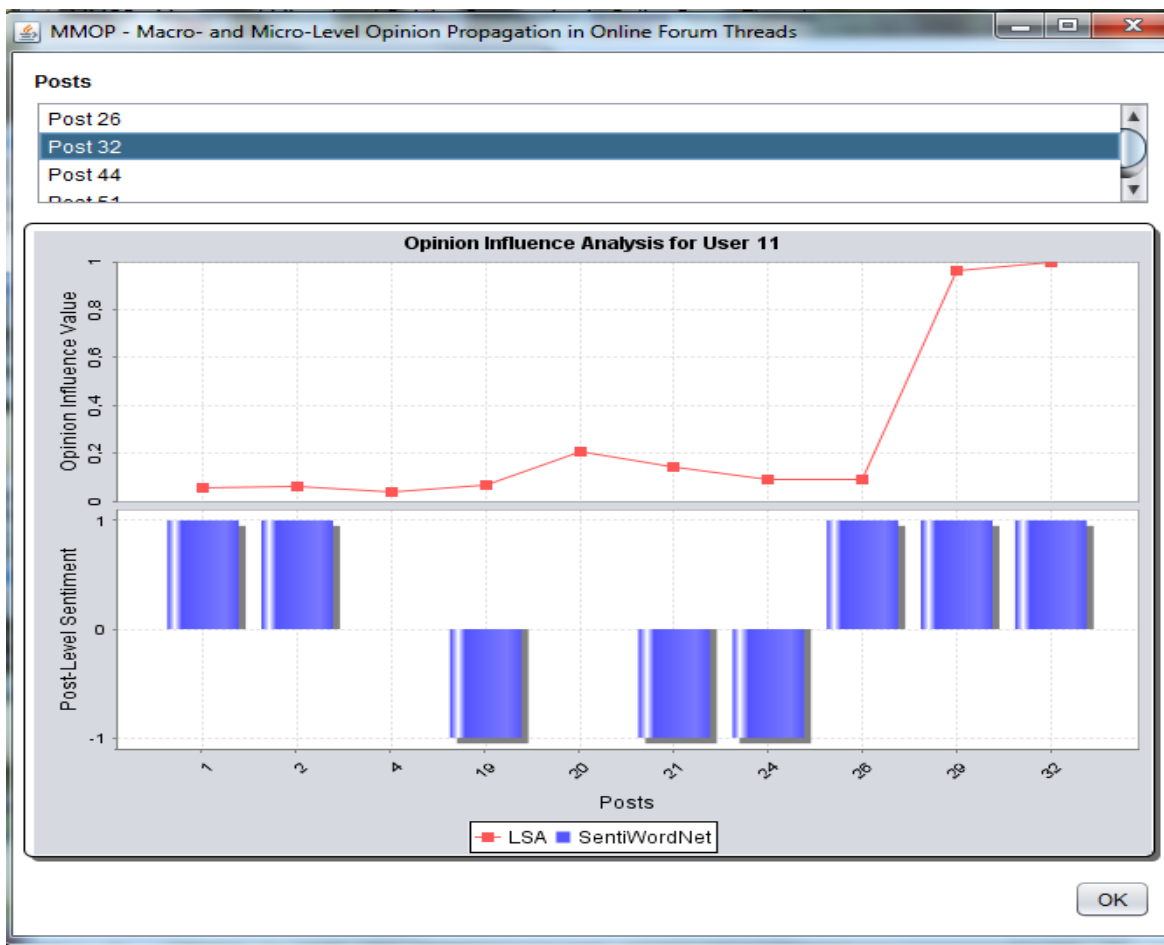


Figura 8.4: Interfața aplicației care permite analiza influenței opiniei Participantului 11, pentru Postarea 32.

CAPITOLUL 9

Concluzii și Direcții Ulterioare

Detectarea propagării opiniilor atât în rețelele de socializare online, cât și în firele de discuție din forumuri este o temă de cercetare complexă și de foarte mare interes în prezent. În această teză, am analizat propagarea opiniilor în mediile sociale urmând trei direcții de cercetare: propagarea opiniilor în firele de discuție din forumuri, propagarea opiniilor în rețelele de socializare online și minierul opiniilor din texte.

Principalele contribuții științifice ale acestei teze vizează prima direcție de cercetare, și anume propagarea opiniilor în firele de discuție din forumuri, care se deosebește fundamental de studiile anterioare axate pe cercetarea propagării opiniilor în rețelele de socializare online. Din cunoștințele noastre, această teză este primul studiu sistematic în problematica propagării opiniilor în forumurile de discuții. În acest scop, am propus trei abordări.

Prima abordare propusă de noi (analiza propagării opiniilor la nivel de utilizator în firele de discuție din forumuri) urmărește să stabilească dacă participanții la discuție fie vor ajunge la un acord total sau parțial, fie, din contră, vor continua să aibă opinii diferite.

A doua abordare pe care noi o propunem (analiza propagării opiniilor la nivel de postare în firele de discuție din forumuri) permite vizualizarea, într-un mod simplificat, a grafului asociat cu firul de discuție din forum, luând în considerare analiza sentimentului la nivel de postare. Pentru a rezolva această abordare, am propus patru algoritmi, iar rezultatele obținute pot fi observate cu ajutorul unor instrumente de vizualizare.

A treia abordare propusă de noi (analiza influenței opiniilor în firele de discuție din forumuri) arată în ce condiții un utilizator este considerat influențat de opiniile exprimate de ceilalți participanți la discuție. În metoda propusă de noi, influența dintre participanți este cuantificată la nivelul opiniilor exprimate de aceștia. Concret, un utilizator este influențat de un alt utilizator / alți utilizatori dacă își schimbă opinia inițială de-a lungul timpului. Această schimbare este consecința interacțiunilor dintre participanții la discuție. În cazul firelor de discuție din forumuri, prin interacțiuni înțelegem postările trimise de un utilizator ca răspuns la postarea anterioară / postările anterioare.

A doua direcție de cercetare a acestei teze este reprezentată de propagarea opiniilor în rețelele de socializare online. În primul rând, vom prezenta stadiul actual al cercetării teoretice în domeniul propagării de opinii în rețelele de socializare online. Astfel, vom analiza cele mai importante modele propuse în literatura de specialitate. Aceste modele prezintă mai multe dezavantaje deoarece nu acoperă toate situațiile care apar într-o rețea de socializare online.

În analiza propagării opiniilor în mediile sociale, o provocare este reprezentată de modul în care opiniile sunt extrase din mesajele / postările schimbate între utilizatori. Metoda bazată pe analiza relațiilor de dependențe între cuvinte nu se aplică pentru unele medii sociale, precum Twitter, unde mesajele conțin greșeli gramaticale, așa cum se arată în studiul nostru de caz din secțiunea 5.4.

A treia direcție de cercetare tratată în teză este reprezentată de procesul mineritului opiniilor. Mineritul opiniilor implică sarcini specifice: identificarea cuvintelor de opinie, estimarea polarității opiniilor, precum și prelucrarea ulterioară a opiniilor obținute (de exemplu, sumarizarea opiniilor sau analiza propagării de opinii). Diverse metode pentru mineritul opiniilor au fost dezvoltate cu scopul de a interpreta multitudinea de date electronice aflate într-o continuă creștere. Mineritul opiniilor este un domeniu deschis cercetării viitoare deoarece există încă multe sarcini de prelucrare a limbajului natural rezolvate doar parțial. Despre toate aceste provocări, noi am discutat în capitolul 5.

9.1. Contribuții Personale

În această secțiune, vom rezuma principalele contribuții ale lucrării:

- Am analizat în detaliu conceptul de *opinie*, subliniind diferențele față de alte concepte conexe.
- Am analizat proprietățile rețelelor de socializare cu un conținut dinamic. Aceste proprietăți explică modul în care structura rețelelor dinamice de socializare se schimbă în timp. Prin urmare, analiza propagării opiniilor în rețelele de socializare poate să țină cont de aceste proprietăți.
- Am analizat provocările apărute în cercetarea mineritului opiniilor din perspectiva prelucrării limbajului natural în vederea creării unui sistem cu o mai bună precizie a rezultatelor. Aceste provocări reprezintă sarcini de prelucrare a limbajului natural: tokenizarea, etichetarea cu părți de vorbire a cuvintelor, parsarea relațiilor de dependență dintre cuvinte, rezolvarea coreferințelor, modificatorii de valență, dezambiguizarea sensurilor cuvintelor de opinii, lematizarea.
- Am analizat patru vocabulare de opinii (SentiWordNet, Micro-WNOp, vocabularul de subiectivitate MPQA, vocabularul de opinii dezvoltat de Bing Liu). Am folosit aceste vocabulare de opinii în experimentele noastre pentru a determina sentimentul (pozitiv, negativ sau neutru) cuvintelor de opinie.
- Am implementat un algoritm existent pentru etichetarea cu partea de vorbire corespunzătoare a cuvintelor din postări. Testele au fost făcute în funcție de doi parametri: variația în dimensiune a corpusului de antrenare, precum și corpusurile de antrenare și de testare au fost din domenii diferite (texte literare și texte din medii colaborative online precum chaturi, wiki-uri, forumuri și Twitter). Din punct de vedere al mineritului opiniilor, principala concluzie care se desprinde din aceste experimente este aceea că părțile de vorbire asociate cuvintelor din postările scrise în firele de forumuri pot fi detectate cu mare precizie. Prin urmare, o analiză bazată pe identificarea relațiilor de dependență dintre cuvintele din postări poate fi utilizată pentru extragerea opiniilor din firele de discuție din forumuri.
- Am descris problema *propagării opiniilor* în rețelele de socializare online.
- Am comparat propagarea opiniilor cu alte fenomene de răspândire întâlnite în rețelele de socializare online, cum ar fi: zvonurile, informațiile, inovațiile tehnologice.

- Am făcut o analiză a conceptelor legate de problema propagării opiniilor în rețelele de socializare online, precum: *influența socială*, *analiza legăturilor sociale*, *analiza rolurilor utilizatorilor* în timpul procesului de propagare de opinii.
- Am discutat șase modele de propagare de opinii în rețelele de socializare online (modelul Voter, modelul Sznajd, modelul Deffuant, modelul Hegselmann-Krause, modelul Independent Cascade, modelul Linear Threshold) și am analizat dezavantajele acestor modele.
- Am introdus direcția de cercetare a propagării opiniilor în firele de discuție din forumuri.
- Am subliniat diferențele dintre procesul de propagare de opinii în rețelele de socializare online (cum ar fi Facebook sau Twitter) și cel din firele de forumuri.
- Am propus o metodă pentru a analiza propagarea opiniilor în firele de discuție din forumuri. Analiza propagării opiniilor se face la nivel de utilizator. În acest scop, am formalizat mai multe concepte. Această metodă stabilește dacă, la un moment de timp determinat, participanții la discuție fie ajung la un acord total sau parțial, fie își mențin opiniile, care pot fi diferite sau chiar contrare.
- Am propus și o altă metodă pentru a analiza propagarea opiniilor în firele de discuție din forumuri. Analiza propagării opiniilor se face la nivel de postare. Această metodă ne permite să urmărim evoluția fluxului de opinii într-un fir de discuție din forum. În acest scop, am creat patru algoritmi.
- Am propus o formulă pentru identificarea sentimentului la nivel de postare în care mai multe opinii de polarități diferite sunt exprimate. În această formulă am luat în considerare tăria sentimentelor exprimate de opinii.
- Am propus o metodă pentru a analiza influența dintre participanții care și-au exprimat opinii în firele de discuție din forumuri.
- Am implementat cele trei metode propuse de noi pentru a studia problema propagării opiniilor în firele de discuție din forumuri, combinând concepte din mai multe domenii: mineritul opiniilor, prelucrarea limbajului natural, teoria matematică a grafurilor, sociologie.
- Am testat cele trei metode propuse de noi pentru analiza propagării opiniilor în fire de discuții reale din forumuri, spre deosebire de metodele propuse în studiile anterioare, care au fost testate pe rețele generate folosind diferite modele de grafuri ce simulează rețelele de socializare online ([8], [41], [7], [5]).

9.2 . Direcții Ulterioare de Cercetare

Analiza propagării opiniilor în mediile sociale are aplicații numeroase în lumea reală, dar este nevoie de mai multă cercetare în acest domeniu pentru a obține rezultate cât mai pertinente. Această teză deschide o direcție nouă cu privire la propagarea opiniilor în firele de discuție din forumuri. În viitor ne propunem să continuăm studiul acestei probleme din două perspective:

- Dorim să continuăm examinarea problemei propagării opiniilor în diferite tipuri de medii sociale, cum ar fi blogurile, forumurile și rețelele de socializare online pentru a descrie asemănările / diferențele dintre procesele propagării opiniilor în aceste medii. În ceea ce privește rețelele de socializare online, modul în care structura lor influențează procesul propagării opiniilor constituie o temă de cercetare pe care o avem în vedere.
- Rezultatele metodelor propuse pentru problema propagării opiniilor în firele de discuție din forumuri depind de tehnicile folosite pentru extragerea opiniilor din postări. Prin urmare, este necesară o îmbunătățire a tehnicilor folosite în mineritul opiniilor. În viitor dorim să luăm în considerare și detectarea opiniilor implicite. De asemenea, sunt necesare metode complexe pentru a clasifica cuvintele de opinie în cele trei categorii de sentimente (pozitive, negative sau neutre), ținând cont de contextul în care acestea apar. Mai mult decât atât, noi dorim să îmbunătățim algoritmul ce identifică sentimentul unei postări în care sunt exprimate opinii de diferite sentimente.

Propagarea opiniilor în mediile sociale rămâne o direcție deschisă de cercetare. O particularitate a acestei direcții de cercetare este dată de faptul că teorii științifice independente pot fi conectate între ele în vederea detectării cu exactitate a propagării opiniilor în mediile sociale.

LISTA DE PUBLICAȚII

1. Cercel, D.-C., & Trăușan-Matu, Ș. (accepted). User-Level Opinion Propagation Analysis in Discussion Forum Threads. In *Proceedings of the 16th International Conference on Artificial Intelligence: Methodology, Systems, Applications (AIMSA 2014)*. Varna, Bulgaria: Springer/LNAI.
(ISI indexed, Nomination for AIMSA'14 Best Paper Award)
2. Cercel, D.-C., & Trăușan-Matu, Ș. (2014). Opinion Propagation in Online Social Networks: A Survey. In *Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS'14)* (pp. 1-10). Thessaloniki, Greece: ACM.
3. Cercel, D.-C., & Trăușan-Matu, Ș. (submitted). Opinion Influence Analysis in Online Forum Threads. In *Proceedings of the 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2014)*. Timișoara, Romania: IEEE.
(ISI indexed, Category C in ARC Conference Ranking)
4. Cercel, D.-C., & Trăușan-Matu, Ș. (submitted). Research Challenges in Opinion Mining from a Natural Language Processing Perspective. In *Scientific Bulletin, University POLITEHNICA of Bucharest, Series C*.
(CNCSIS B+)
5. Cercel, D.-C., & Trăușan-Matu, Ș. (2013). A POS Tagger analysed in collaboration environments and literary texts. In *Proceedings of the 13th Francophone International Conference on Knowledge Discovery and Management (EGC'2013)* (vol. RNTI-E-24, pp. 287-292). Toulouse, France.
(Category C in ARC Conference Ranking)
6. Cercel, D.-C., & Trăușan-Matu, Ș. (2013). The Features-Based Evolution of the Opinions in Forums Using a Semantic Summarization Approach. In *Proceedings of the Second International Workshop on Semantic and Collaborative Technologies for the Web, in conjunction with the 2nd International Conference on Systems and Computer Science (ICSCS 2013)* (pp. 222–227). Villeneuve d'Ascq, France: IEEE.
7. Cercel, D.-C., Trăușan-Matu, Ș., & Velcin, J. (2013). Overview and Preliminary Results for an Opinion Summarization Method. In *Proceedings of the 19th International Conference on Control Systems and Computer Science (CSCS19)* (pp. 218-224). Bucharest, Romania: IEEE.
(ISI indexed)
8. Cercel, D.-C., & Trăușan-Matu, Ș. (2013). Evoluția opiniilor în forum-uri folosind o abordare semantică de sumarizare. In *A 10-a Conferința Națională de Interacțiune Om-Calculator (RoCHI 2013)* (pp. 51-56). Cluj-Napoca, Romania.
(CNCSIS B+)
9. Cercel, D.-C., & Trăușan-Matu, Ș. (2012). POS tagger bazat pe modelul HMM de ordinul doi. In *Revista Română de Interacțiune Om-Calculator*, 5(2), 15-34. (CNCSIS B+)

REFERINȚE

- [1] A. E. S. Baccianella and F. Sebastiani, "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," in *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, 2010.
- [2] S. Cerini, V. Compagnoni, A. Demontis, M. Formentelli, and G. Gandini, "Micro-WNOp: A gold standard for the evaluation of automatically compiled lexical resources for opinion mining," in *Language resources and linguistic theory: Typology, second language acquisition, English linguistics*, A. Sans, Ed., ed: Franco Angeli Editore, 2007.
- [3] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity in phrase-level sentiment analysis," presented at the Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, British Columbia, Canada, 2005.
- [4] M. Hu and B. Liu, "Mining and summarizing customer reviews," presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA, 2004.
- [5] K. Suchecki, V. M. Eguíluz, and M. San Miguel, "Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution," *Physical Review E*, vol. 72, p. 036132, 2005.
- [6] J. Sznajd and K. Sznajd-Weron, "Opinion evolution in closed community," *International Journal of Modern Physics C*, vol. 11, pp. 1157-1165, 2000.
- [7] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, "Mixing beliefs among interacting agents," *Advances in Complex Systems*, vol. 3, pp. 87-98, 2000.
- [8] R. Hegselmann and U. Krause, "Opinion Dynamics and Bounded Confidence, Models, Analysis and Simulation," *Journal of Artificial Societies and Social Simulation*, vol. 5, p. 2, 2002.
- [9] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," presented at the Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, Washington, D.C., 2003.
- [10] B. Liu, *Web Data Mining Exploring Hyperlinks, Contents and Usage Data*: Springer-Verlag 2011.
- [11] S.-M. Kim and E. Hovy, "Determining the sentiment of opinions," presented at the Proceedings of the 20th international conference on Computational Linguistics, Geneva, Switzerland, 2004.
- [12] X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion mining," presented at the Proceedings of the 2008 International Conference on Web Search and Data Mining, Palo Alto, California, USA, 2008.
- [13] B. Liu, *Sentiment Analysis and Opinion Mining*: Morgan & Claypool Publishers, 2012.
- [14] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification," presented at the Proceedings of the 14th ACM international conference on Information and knowledge management, Bremen, Germany, 2005.
- [15] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," presented at the Proceedings of the ACL-02 conference on Empirical methods in natural language processing, 2002.
- [16] P. D. Turney and M. L. Littman, "Measuring praise and criticism: Inference of semantic orientation from association," *ACM Trans. Inf. Syst.*, vol. 21, pp. 315-346, 2003.
- [17] X. F. Wang, "Complex Networks: Topology, Dynamics and Synchronization," *International Journal of Bifurcation and Chaos*, vol. 12, pp. 885-916, 2002.
- [18] M. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, vol. 45, pp. 167-256, 2003.
- [19] S. Wasserman and K. Faust, *Social network analysis: methods and applications*. Cambridge; New York: Cambridge University Press, 1994.
- [20] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," presented at the Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, San Diego, California, USA, 2007.
- [21] A. Java, X. Song, T. Finin, and B. Tseng, "Why We Twitter: An Analysis of a Microblogging Community," in *Advances in Web Mining and Web Usage Analysis*. vol. 5439, H. Zhang, M.

- Spiliopoulou, B. Mobasher, C. L. Giles, A. McCallum, O. Nasraoui, *et al.*, Eds., ed: Springer Berlin Heidelberg, 2009, pp. 118-138.
- [22] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, pp. 440-442, 1998.
- [23] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, vol. 286, pp. 509-512, 1999.
- [24] L. Hamill and N. Gilbert, "Simulating Large Social Networks In Agent-Based Models: A Social Circle Model," *Emergence: Complexity & Organization*, vol. 12, 2010.
- [25] A. Alaali, M. A. Purvis, and B. T. R. Savarimuthu, "Vector Opinion Dynamics: An Extended Model for Consensus in Social Networks," in *Web Intelligence and Intelligent Agent Technology*, 2008, pp. 394-397.
- [26] S. Fortunato, V. Latora, A. Pluchino, and A. Rapisarda, "Vector Opinion Dynamics in a Bounded Confidence Consensus Model," *International Journal of Modern Physics C*, vol. 16, pp. 1535-1551, 2005.
- [27] C. D. Manning and H. Schütze, *Foundations of statistical natural language processing*: MIT Press, 1999.
- [28] M.-C. de Marneffe and C. Manning, *Stanford typed dependencies manual*, 2011.
- [29] V. Stoyanov, C. Cardie, N. Gilbert, E. Riloff, D. Buttler, and D. Hysom, "Coreference resolution with reconcile," presented at the Proceedings of the ACL 2010 Conference Short Papers, Uppsala, Sweden, 2010.
- [30] X. Ding and B. Liu, "Resolving object and attribute coreference in opinion mining," presented at the Proceedings of the 23rd International Conference on Computational Linguistics, Beijing, China, 2010.
- [31] L. Polanyi and A. Zaenen, "Contextual Valence Shifters," in *Computing Attitude and Affect in Text: Theory and Applications*. vol. 20, J. Shanahan, Y. Qu, and J. Wiebe, Eds., ed: Springer Netherlands, 2006, pp. 1-10.
- [32] A. Balahur and A. Montoyo, "OpAL: Applying Opinion Mining Techniques for the Disambiguation of Sentiment Ambiguous Adjectives in SemEval-2 Task 18," 2010.
- [33] T. Brants, "TnT: a statistical part-of-speech tagger," presented at the Proceedings of the sixth conference on Applied natural language processing, Seattle, Washington, 2000.
- [34] W. N. Francis and H. Kucera, "Brown Corpus Manual," Department of Linguistics, Brown University, Providence, Rhode Island, US, 1979.
- [35] A. Stavrianou, "Modeling and mining of web discussions," University of Lyon, France, PhD Thesis, 2010.
- [36] R. Zafarani, W. D. Cole, and H. Liu, "Sentiment propagation in social networks: a case study in livejournal," presented at the Proceedings of the Third international conference on Social Computing, Behavioral Modeling, and Prediction, Bethesda, MD, 2010.
- [37] G. Somprasertsri and P. Lalitrojwong, "Mining Feature-Opinion in Online Customer Reviews for Opinion Summarization," 2010.
- [38] M. A. Walker, J. E. F. Tree, P. Anand, R. Abbott, and J. King, "A Corpus for Research on Deliberation and Debate," in *LREC*, 2012, pp. 812-817.
- [39] J. Tang, J. Sun, C. Wang, and Z. Yang, "Social influence analysis in large-scale networks," presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France, 2009.
- [40] X. Shuai, Y. Ding, J. Busemeyer, S. Chen, Y. Sun, and J. Tang, "Modeling Indirect Influence on Twitter," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 8, pp. 20-36, 2012.
- [41] L. da F. Costa and F. A. Rodrigues, "Surviving opinions in Sznajd models on complex networks," *International Journal of Modern Physics C*, vol. 16, pp. 1785-1792, 2005.