Poboljšani BSP algoritam za
analizu socijalnih mreža

**Improved BSP Clustering Algorithm for**

**Social Network Analysis**

Dragan Veljović1

Elektrotehnički fakultet Univerziteta u Beogradu1

**Sadržaj –** Grupisanje u analizi socijalnih mreža se razlikuje od tradicionalnog grupisanja. Zahteva grupisanje objekata u klase na osnovu njihovih veza, kao i atributa. Na osnovu BSP (business system planning -planiranje poslovnih sistema) algoritma, predložen je algoritam za grupisanje u analizi socijalnih mreža. Predloženi algoritam može da grupiše objekte u socijalnim mrežama u različite klase na osnovu njihovih veza i da identifikuje relacije između klasa dinamički i zahteva manju količinu memorije.

**Abstract -** The clustering in social network analysis is different from traditional clustering. It requires grouping objects into classes based on their links as well as their attributes. On the basis of BSP (business system planning) clustering algorithm, a social network clustering analysis algorithm is proposed. The proposed algorithm, can group objects in a social network into differen classes based on their links and identify relation among classes dynamically & require less amount of memory .

1. uvod

Socijalne mreže imaju strukturu grafa gde čvorovi predstavljaju ljude ili druge entitete uključene u socijalni kontekst. One su visoko dinamične, promenjivih relacija između ljudi i drugih entiteta. Dinamičke osobine socijalnih mreža čine njihovo proučavanje izazovnim, upravo se time analiza socijalnih mreža bavi.

Analiza socijalnih mreža je polje istraživanja u sociologiji, koje je u poslednje vreme privuklo puno pažnje u oblasti skrivenog znanja. Sa aspekta pronalaženja skrivenog znanja, socijalna mreža je heterogena, više-relacioni skup podataka predstavljen pomoću grafa.

U ovom dokumentu, kao metodologija iz oblasti skrivenog znanja koristi se grupisanje i to BSP algoritam. Algoritam može da grupiše objekte socijalnih mreža u različite klase na osnovu njihovih linkova. Takođe, može da identifikuje relacije između klasa.

1. Analiza Socijalnih mreža zasnovana na BSP grupisanju

Tradicionalni algoritmi za grupisanje razdvajaju objekte u klase na osnovu njihove sličnosti. Objekti u klasi su slični jedni drugima i veoma različiti od objekata u drugim klasama. S druge strane grupisanje u analizi socijalnih mreža grupiše objekte i na osnovu njihovih veza, što predstavlja najveći izazov.

BSP (business system palnning – planiranje poslovnih sistema) algoritam je predložio IBM. Dizajniran je da definiše informacione strukture u firmama prilikom procesa planiranja poslovnih sistema. Ovaj algoritam grupiše poslovni proces u podsisteme i definiše relacije između njih.

U osnovi BSP algoritam koristi objekte i veze između njih u procesu grupisanja. Obzirom da i u analizi socijalnih mreža postoji postoje objekti odgovarajuće relacije, ovaj algoritam može da se primeni.

Prema teoriji grafova, socijalna mreža je direktan graf gde su čvorovi objekti, a grane relacije tog grafa. Primer tog grafa je prikazan na slici 1.



Slika 1. Primer socijalne mreže

Neka je Oi ( i = 1...m ) objekat u socijalnoj mreži i neka je Ej ( j = 1...n ) direktna veza između dva objekta, tada je:

1. Relacija dostupnosti iz jednog koraka: relacija između Oi i Oj akko uključuje jednu i samo jednu direktnu granu grafa.
2. Relacija dosupnosti iz više koraka: relacija između Oi i Oj akko uključuje dve ili više direktnih grana grafa.

Nakon definisanja dva gore pomenuta termina možemo koristiti BSP algoritam za analizu socijalnih mreža. Ta analiza prolazi kroz sledeće korake:

**Generisanje matrice početka veza i matrice kraja veza.**

Definišu se dve matrice: Lc i Lp. Lc je matrica početka veze (grane grafa) – Lc(i, j) = 1 ako se čvor grafa Oi nalazi na repu veze Ej istog grafa. Lp je matrica kraja veze (grane grafa) – Lp(i, j) = 1 ako se čvor grafa Oi nalazi na vrhu veze Ej istog grafa.

**Računanje matrice dostupnosti iz jednog koraka**

Nakon određivanja matrica Lc i Lp, matrica dostupnosti iz jednog koraka između objekata se računa po formuli:

 (1)

Gde je ^ Bulov proizvod, a V Bulova suma.

G(i, j) = 1 znači da od Oi do Oj relacija dostupnosti iz jednog koraka, s druge strane G(i, j) = 0 znači da takva relacija ne postoji.

**Računanje matrice dostupnosti iz više koraka**

Računanje se vrši na osnovu sledećih jednačina:

 , (2)

Ove matrice uključuju relacije dostupnosti iz dva, tri, ..., m-1 koraka.

**Računanje matrice opšte dostupnosti.**

Obzirom da je bitno da li postoji relacija izmđu dva objekt, a ne da li ona ide preko dva, tri, ili više objekata, matrica opšte dostupnosti se računa kao:

 , (3)

Gde je V Bulova suma, a I jedinična matrica.

R(i, j) = 1 znači da postoji relacija između Oi i Oj, ali relacija nije obostrana, relacija između Oj i Oi ne mora da postoji.

**Računanje matrice zajedničke dostupnosti**

Matrica zajedničke dostupnosti se računa koristeći jednačinu:

 (4)

Gde ^ predstavlja Bulov proizvod.

U matrici Q(i, j) = 1znači da postoji uzajamna relacija između Oi i Oj.

U socijalnim mrežama ako dva objekta imaju relaciju zajedničke dostižnosti, onda oni pripadaju istoj klasi, odnosno klasteru.

**Identifikacija relacija između klasa.**

Nakon grupisanja neophodno je identifikovati veze između klasa. Ovo može da se uradi formiranjem matrice dostupnosti iz jednog koraka za generisane klase (matrixG).

Kompletan algoritam u pseudokodu:

**Input:**

Lc : matrica početka veze

 Lp: matrica kraja veze

**Begin**

  *C*

**End**

*C* , predstavlja generisanje klasa na osnovu matrice zajedničke dostižnosti. , znači identifikovanje relacija između klastera na osnovu matrice dostižnosti iz jednog koraka.

1. PredloženO POBOLJŠANJE

Glavni nedostatak gore pomnutog algoritma je što koristi matrice. U realnoj analizi dimenzije ovih matrica mogu biti izuzetno velike, s druge strane broj nenultih elemenata nije procentualno veliki. Stoga je predložena sledeća struktura za čuvanje ovih elemenata u formi ulančane liste:

Struct node {

 int row, col, val;

 node\* next;

}

Na osnovu slike 1. dat je sledeći primer:



Slika 2. Lc i Lp – reprezentacija u formi ulančane liste

**Računanje matrice dostupnosti iz jednog koraka**

****

Slika 3. Matrica dostupnosti iz jednog koraka u formi ulančane liste

**Računanje matrice zajedničke dostupnosti**

****

Slika 4. Matrica zajedničke dopunosti

Na osnovu Q, graf sa slike 1. se može podeliti u dve klase C1(uključuje objekte O1, O2, O3) i C2(uključuje objekte O4, O5, O6).

**Identifikacija relacija između klasa**

Na osnovu matrice dostupnosti iz jednog koraka G postoji dostuonost iz jednog koraka između O2 i O4, O3 i O4 tako da između klastera postoji povezanost prikazana na slici 5.



Slika 5. Rezultat grupisanja.

1. zaključak

Na osnovu BSP algoritma za grupisanje, predložen je algoritam za grupisanje u socijalnim mrežama. On deli socijalnu mrežu na klase u zavisnosti od objekata u socijalnim mrežama i veza između njih. Takođe može da identifikuje relacije između klasa.

U datoj analizi pošlo se od pretpostavke da sve veze imaju istu težinu. Međutim, u stvarnom svetu situacija nije takva, veze mogu biti različitih težina.

Osobine pojedinačnih grupa nisu uzete u razmatranje.

To su mesta gde treba tražiti buduća poboljšanja i u tom pravcu treba usmeriti dalje istraživanje s ciljem poboljšanja algoritma.

literatura

1. <http://kelasti.files.wordpress.com/2009/05/bsp2b.pdf>, 30.12.2014.
2. Han J, Kamber M., “**Data Mining: Concepts and Techniques**”, 2nd edition, The Morgan Kaufmann Publishers,San Franciso, 2006
3. Milutinović, V., “**The Best Method for Presentation of Research Results**”, IEEE TCCA Newsletter, September 1997, pp. 1-6