

Merenje sličnosti grafova i njihovih čvorova uparivanjem suseda

Mladen Nikolić

20. novembar, 2010.

Overview

- 1 Uvod
- 2 Osnovni pojmovi
- 3 Računanje sličnosti čvorova uparivanjem suseda
- 4 Eskperimentalna evaluacija
- 5 Zaključci

Overview

- 1 Uvod
- 2 Osnovni pojmovi
- 3 Računanje sličnosti čvorova uparivanjem suseda
- 4 Eskperimentalna evaluacija
- 5 Zaključci

Šta je slično?

- Zanima nas topološka sličnost
- Izomorfizam?
- Nešto “labavije”?
- Nema privilegovane definicije

Pristupi

- Sličnost grafova
 - Sličnost raspodela različitih statistika
 - Veličina najvećeg zajedničkog podgrafa ili najmanjeg zajedničkog nadgrafa
 - ...
- Sličnost čvorova grafova
 - Dva čvora su utoliko sličnija sto su sličniji njihovi susedi

Primer metode za računanje sličnosti čvorova

$$x_{ij}^0 \leftarrow 1$$
$$x_{ij}^{k+1} \leftarrow \sum_{(p,i) \in E_A, (q,j) \in E_B} x_{pq}^k + \sum_{(i,p) \in E_A, (j,q) \in E_B} x_{pq}^k$$
$$X^{k+1} \leftarrow \frac{X^{k+1}}{\|X^{k+1}\|_2}$$

Poželjna svojstva koja nedostaju postojećim metodama

- Kada se graf poredi sa samim sobom, svaki čvor treba da bude najbližnji sebi
- Raspon skorova sličnosti treba da bude poznat
- Sličnost čvora sa samim sobom treba da bude maksimalna
- Sličnost čvorova treba da ima “apsolutno” značenje (nasuprot relativnom)
- Čvorove kojima nedostaju ulazne/izlazne grane treba smatrati sličnim upravo zbog nedostatka

Overview

- 1 Uvod
- 2 Osnovni pojmovi**
- 3 Računanje sličnosti čvorova uparivanjem suseda
- 4 Eskperimentalna evaluacija
- 5 Zaključci

Problem pridruživanja (eng. assignment problem)

- Dva skupa objekata A i B
- Težinska funkcija $w(a, b)$
- Uparivanje $M = \{(i, j) | i \in A, j \in B\}$
- Funkcije enumeracije $M = \{(f(l), g(l)) | l = 1, 2, \dots, k\}$
- Problem pridruživanja se sastoji u nalaženju uparivanja maksimalne težine $\sum_{l=1}^k w(f(l), g(l))$
- Mađarski algoritam (1955)

Overview

- 1 Uvod
- 2 Osnovni pojmovi
- 3 Računanje sličnosti čvorova uparivanjem suseda**
- 4 Eskperimentalna evaluacija
- 5 Zaključci

Nov princip

- Dva čvora i i j se smatraju sličnim ako se susedi čvora i mogu upariti sa sličnim susedima čvora j

Računanje sličnosti

$$x_{ij}^0 \leftarrow 1$$

$$x_{ij}^{k+1} \leftarrow \frac{s_{in}^{k+1}(i,j) + s_{out}^{k+1}(i,j)}{2}.$$

$$s_{in}^{k+1}(i,j) \leftarrow \frac{1}{m_{in}} \sum_{l=1}^{n_{in}} x_{f_{ij}^{in}(l)}^k g_{ij}^{in}(l)$$

$$s_{out}^{k+1}(i,j) \leftarrow \frac{1}{m_{out}} \sum_{l=1}^{n_{out}} x_{f_{ij}^{out}(l)}^k g_{ij}^{out}(l)$$

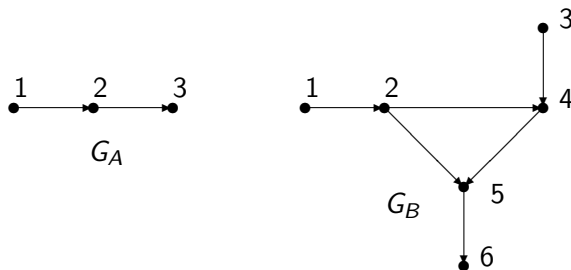
$$m_{in} = \max(id(i), id(j))$$

$$m_{out} = \max(od(i), od(j))$$

$$n_{in} = \min(id(i), id(j))$$

$$n_{out} = \min(od(i), od(j))$$

Primer



| | 1_B | 2_B | 3_B | 4_B | 5_B | 6_B |
|-------|-------|-------|-------|-------|-------|-------|
| 1_A | 0.682 | 0.100 | 0.597 | 0.200 | 0.000 | 0.000 |
| 2_A | 0.000 | 0.364 | 0.045 | 0.195 | 0.400 | 0.000 |
| 3_A | 0.000 | 0.000 | 0.000 | 0.091 | 0.091 | 0.700 |

Teorijski rezultati

- Za bilo koja dva grafa skorovi sličnosti čvorova konvergiraju kako se broj iteracija povećava
- Skorovi sličnosti su u intervalu $[0, 1]$
- Za dva izomorfna grafova, čvor jednog grafa i njegova slika pri izomorfizmu imaju sličnost 1
- Sličnost svakog čvora sa samim sobom je 1
- Metoda ima sva svojstva nabrojana na početku

Sličnost celih grafova

- Koristeći matricu sličnosti kao težinsku funkciju, rešiti problem pridruživanja između čvorova dva grafa

$$s(G_A, G_B) = \frac{1}{n} \sum_{l=1}^n x_{f(l)g(l)}$$

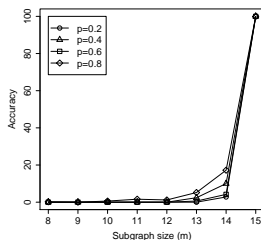
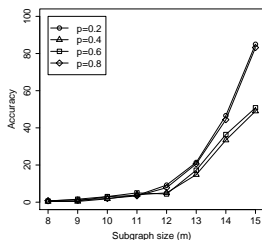
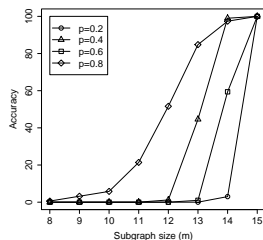
Overview

- 1 Uvod
- 2 Osnovni pojmovi
- 3 Računanje sličnosti čvorova uparivanjem suseda
- 4 Eskperimentalna evaluacija**
- 5 Zaključci

Pronalaženje izomornog podgrafa

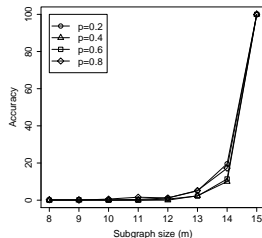
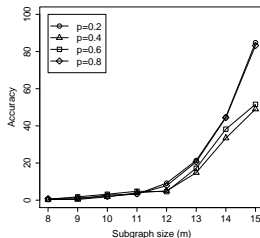
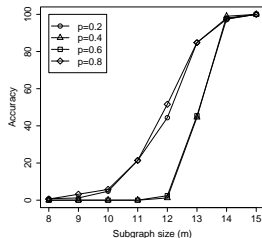
- Problem je pronalaženje podgrafa grafa A koji je izomorfan datom grafu B
- Pristup pomoću mera sličnosti čvorova
 - Izračunati skorove sličnosti
 - Naći uparivanje najveće težine za čvorove grafova A i B
 - Definisati izomorfizam na osnovu dobijenog uparivanja

Eksperimentalni rezultati



| | NM | HS | ZV |
|------------|-------|--------|------|
| Preciznost | 27.3 | 17.5 | 13.9 |
| Vreme | 2062s | 11511s | 349s |

Eksperimentalni rezultati



| | NM* | HS* | ZV* |
|------------|------|--------|------|
| Preciznost | 37.8 | 17.5 | 15.0 |
| Vreme | 838s | 11730s | 230s |

Klasifikacija iskaznih formula

- Iskazne formule su grupisane u klase na osnovu njihovog porekla
- Iskazne formule imaju grafovsku strukturu
- Mere sličnosti nad grafovima se mogu upotrebiti za klasifikaciju grafova pomoću algoritma k-najbližih suseda.
- Preciznost klasifikacije pomoću grafovske sličnosti je 93%
- Preciznost pristupa specijalno konstruisanog za ovu svrhu je 96%

Overview

- 1 Uvod
- 2 Osnovni pojmovi
- 3 Računanje sličnosti čvorova uparivanjem suseda
- 4 Eskperimentalna evaluacija
- 5 Zaključci**

Zaključci

- Nova formulacija principa sličnosti
- Nova metoda
- Dokazana konvergencija i neka bitna svojstva koja druge metode nemaju
- Na problemu pronalaženja izomornog podgraфа nova metoda daje bolje rezultate nego postojeće
- Postoje domeni u kojima je sličnost definisana novom metodom smisljena