

Tisztelt Dr. Kovács László!

Nagyon szépen köszönöm a dolgozatom nagyon alapos vizsgálatát. Az észrevételeit illetve a kérdését az alábbiakban válaszolom meg.

1. Az alkalmazott jelölések nem minden esetben definiáltak az első felhasználás helyén. Például a 25. oldalon  $G_{\{i,j\}}$  értelmezésében az éltípusok csak egy része definiált, mások nem  $(o, m)$ .

Nagyon szépen köszönöm az észrevételt. Valóban ezek nincsenek korábban feltüntetve, mint a 2.1-es táblázat, azonban az ezeket részletesen tartalmazza. Igen, illő lett volna, itt is ezeket feltüntetni.

- 
2. A 25. oldalon bemutatott többrétegű hálózat esetében kiemelésre került, hogy "the interlayer connections help in term of visualization (Figure 2.2)". A szintek közötti kapcsolat viszont nem kapott pontos definíciót, az ábrában sem jelenik meg explicite. Ha jól értettem, a kapcsolat azt jelenti, hogy egy adott csomópont minden szinten azonos koordinátában jelenik meg. Miért nevezhető ez kapcsolati elemnek?

Ezek speciális kapcsolati elemek, a multiplex hálózat definíciója szerint, az említett elemek a szupraadjecencia mátrix identikus mátrixban lévő identikus éleket mutat. Ez ábrázolva úgy látható, hogy az ugyanolyan csomópontok egy vonalba kerülnek a szinttől függetlenül.

- 
3. A 27 oldalon a Table 2.1 táblában a P sorában a méret nem pontos ( $N_p \times N_m$  helyett  $N_p \times N_p$ )

Valóban itt  $N_p \times N_m$  méretű mátrix a helyes.

- 
4. A Table 2.2 (28. oldal) esetében furcsának hat, hogy a "Definition" kifejezést definiáljuk, mint "Material, energy and information flow"-t.

Igen valóban furcsán hat, a definíció szó használata itt. Itt a meghatározás és leírás jelentését szerettem volna használni a szónak, viszont ezt valóban csak a cím elolvasásával együtt jelenik talán meg. De egyet értek, nem ilyen alkalomra szokták tartogatni a definition-t.

- 
5. A 34. oldalon szereplő  $\text{sim}()$  hasonlóság definícióban a  $(1,2)$  és  $(1,3)$  halmazok hasonlósága  $\frac{1}{3}$ . Mi indokolta ezen mérőszám választását? Mennyivel jobb ez a változat, mint egy "overlap metric"?

Jaccard szimilitást az érthetősége és értelmezhetősége miatt választottam. Például a termékek és eszközök vonatkoztatásában a overlap az eszközökre vonatkoztatott mérőszám

lenne, az operátorok és állomások vonatkoztatásában pedig az állomásokra vonatkoztatott mérőszám lenne. Ezért ezt minden esetben újra kéne értelmezni. A jaccard szimilaritást használva, mindig ugyanaz marad a viszonyítás, a két halmaz uniója.

$$\text{overlap}(A, B) = \left( \frac{|A \cup B|}{\min(|A|, |B|)} \right), \text{jaccard}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

---

6. A (2.6) és (2.7) képleteknél az eredmény mátrix mérete és indexelése nem tűnik helyesnek, mivel a transzponálás során felcserélődik a sor/oszlop szerep. Nincs összhangban a (2.1) jelölési módszerével.

Valójában ezek helyesek, pont azért mert felcserélődnek a sor/oszlop szerepek.

---

7. A 36. oldalon a (2.9) képletben szereplő szimbólumok között van nem definiált is. Például a képletben  $L$  szerepel a szövegben  $L_c$ .

Cluster modularityben a  $L$  az összes linket jelöli az  $L_c$  a  $c$  indexű community-be tartozó linkeket jelöli. Valóban illő lett volna az  $L$ -t is bevezetni.

---

8. A 60. oldalon a Q modularity metric az  $L$  változó "L represents the number of edges" értelmezésű. Az irodalomban viszont "L total number of stubs" van, ami egy eltérő mennyiség.

A Barabási Albert László féle networksciencebook-ot (<http://networksciencebook.com/>) használtam a jelölések alapjának. A stubs kifejezést nem ismerem, ilyen irodalmi jelöléssel én nem találkoztam.

---

9. Egy mellékes megjegyzés ahhoz, hogy 'LD are forming Semantic Web' (79 oldal): A szemantikus webet tágabban szokás értelmezni, több réteget is tartalmaz, például beleértve a logikát és a következtető motort.

Igen, itt arra szerettem volna hivatkozni, hogy minden leírási mód a szemantikus weben az LD formában van.

---

10. A 83.-oldaltól kezdődő részben néhány pont kevésbé érthető a külső olvasó számára. A E él első definíciója a multi-dimenzionális hálóban  $(u,v,d)$  ahol  $d$  egy tetszőleges címke (dimenzió); majd pár sorral lentebb  $(u,v,d_u,d_v,d_e)$ , ahol már előírjuk 3 eltérő szerepű címke meglétét, majd pár sorral később  $(u,v,D_a)$  lesz az értelmezés, ahol  $a$  tehát a címkék tetszőleges eloszlású halmaza szerepel. Azt is jó lenne itt megadni, hogy  $D_e$  és  $D_v$  diszjunkt-e vagy sem. A 84. oldal első formulájánál is hasznos lenne egy pontosítás. A  $(u,v,D_a)$  értelmezésnél nem számít, hogy a címke mely csomópontokra vonatkozik. Miért nem célszerű közvetlenül a csomópontokhoz rendelni a  $D_v$  címkéket?

Az adatok leírásából / reprezentációjáról származik, ez a leírásmód. A linked data-ban a reprezentáció a 4.1 ábrának a teteje mutatja, Linked Data dataset alcímmel. Hogy miket tartalmaznak ezek a halmazok az a módszer szempontjából elhanyagolható az értelmezésnél lesz kritikus. Az élek dimenziói a  $d_e$  általában diszjunkt a csomópont halmazaitól,  $d_u$ ,  $d_v$  vagyis a

$D_v$ -től. De ez nem egy abszolút kikötés. Erre nincsen definíció a LD-ban.

Az  $\alpha$  jelölés már egy szintet jelöl, vagyis szimultán matchelő dimenziókat. Ez szemléltetve a 4.1 ábrának a jobb felső sarkában látszik. Közvetlenül a csomópontokhoz vannak rendelve a címkék, azonban az link data leírás módja szerint. A kezdetekben máshogy ezt nem lehet csinálni vagy megsérülne a LD leírási mód.

---

11. A 85.oldalon szerepel az alábbi állítás: “the set  $F = [a..M]$  holds the frequent itemsets.”. A definíció szerint viszont  $F$  egy egészeket tartalmazó intervallum. Valószínűleg itt az index jelölése maradt ki.

Ezek a szimultán matchelő dimenziókat mutatják. Vagyis él és csomópont dimenziók, nem egészek, hanem értelmező halmazok.

---

12. Hasznos lenne bemutatni a “reachibility” alkalmazott értelmezését és megvalósítási módszerét, a szövegezés alapján egyfajta származtatási, következtetési lépést jelenthet ebben a kontextusban.

A 4.1 ábra bal alsó ábráján mutatom ezt be. A megvalósítása azonban inkább technikai mint érdekes. És technikailag valóban ez egyfajta objektum származtatás, azonban amikor hálózatként nézzük akkor valójában elérhetőséget nézünk.

---

13. A 99. oldalon ez ll: “scalability of the method is not limited, but it is more an engineering challenge than a research objective”. Általánosságban nem értek egyet ezzel a karakteres megállítással, mert a gyakorlatban (márpedig végső soron valahol mindig a végcél), van kapacitáskorlát (például napjaink aktuális kérdése a red-AI problémája szemben a green-AI-val) és az optimális algoritmus megtalálása sokszor kutatási eredmény.

Igen, abszolút nem volt szándékomban az algoritmus kutatást vagy a hatékonyság / branchelés / stb. másodlagosnak minősíteni.

---

Kérdés a jelölthöz:

Milyen lényegi különbségeket lát a javasolt többdimenziós hálózati modell és a több-attribútumos hálózati adatmodell (Neo4J) között strukturális és funkcionális szinten?

A Neo4J adatbázist nem ismerem a kellő mélységében a technológiáját csak az alapjait és lekérdező nyelvét, ám saját gépen futtatva amikor használni szerettem volna borzasztóan lassú volt a feladathoz. Viszont ez ha jól gondolom akkor nagyon hasonló ahhoz, hogy XML cellát hozzunk létre, vagy egy külön adattáblát. A cél szempontjából lényegtelen, az adatok megmaradnak, viszont feldolgozás szempontjából és használhatóság szempontjából lesz eltérő.

Feldolgozás szempontjából hogy egy JOIN-t vagy egy új lekérdezt indítunk, vagy pedig parsooljuk a XML-t lesz a különbség. Használhatóság szempontjából, hogy csak egy stringet változtatunk vagy egy teljes update-et hozunk létre kulcs validálásokkal lesz a különbség. Ugyanezt érzem a másik esetben is.

Másik amire gondolok, a 10 kérdésben is megemlítem, ha a nyers adatokat nézzük akkor ott még nem lehet közvetlenül csomóponthoz kötni objektumokat, ugyanis az LD leírási módja nem teszi ezt lehetővé.

Nagyon szépen köszönöm!

Tisztelettel,

Honti Gergely Marcell

Honti Gergely  
2021. 12. 22

Ui.: Boldog karácsonyt kívánok Önnek és családjának!