

Opponensi vélemény PhD dolgozat nyilvános védésére

Dolgozat címe: Multidimensional network-based analysis of complex systems

Szerző: Honti Gergely Marcell

Doktori iskola: Pannon Egyetem, Vegyészmérnöki és Anyagtudományok Doktori Iskola

Év: 2021

A dolgozat témaköre egy újszerű többszintű és több-címkés multi-gráf modell bevezetése a komplex rendszerek, elsődlegesen a gyártórendszerek elemzésére.

A témakör választást aktuálisnak és a gyakorlat számára fontosnak tartom. Külön kiemelő, hogy az elvégzett kutatás szervesen kapcsolódik a gazda intézet korábbi doktori kutatási témaköreihöz is, egymásra épülő eredményeket láthatunk az elkészült és megvédett dolgozatokban.

A dolgozat logikailag és szerkezetileg jól felépített, egységes képet mutat, melyben az elméleti modellezés-módszertani eredmények mellett számos gyakorlati példát is tartalmaz a kidolgozott módszerek igazolására. Egyedül a negyedik fejezetnél érzek egy lazább kapcsolódást az előzményekhez.

Külön kiemelő az igen részletes és a széleskörű irodalmi háttérelmzés, melyben a jelölt mintegy 302 irodalmi forrást használ fel. A 302 hivatkozásból mintegy 170 a dolgozat törzsében szerepel, a többi 130 a függelékben megadott leírásokhoz kapcsolódik.

A dolgozat több lényeges tartalmi tisztításon ment át az induló, munkahelyi vitán bemutatott változathoz képest, viszont néhány részletet érintő pontban továbbra is maradtak számomra kérdéses részek, melyek főleg a formális modell leíráshoz kapcsolódnak.

Tartalmi megjegyzések

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

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

5. A 34. oldalon szereplő $\text{sim}()$ hasonlóság definíciójában 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"?

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.

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

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.

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.

10. A 83.-oldaltól kezdődő részben néhány pont kevésbé érthető a külső olvasó számára. A E 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?

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

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

13. A 99. oldalon ez II: "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 gyakorlat 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.

Tézisek értékelése

A jelölt által megadott téziseket a megadott formában elfogadom. A tézisek a jelölt által elért fontosabb eredményeket fogalmazzák meg.

Kérdés a jelölthöz:

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

Összegzés

Az értekezés egy komplex gyakorlati probléma megoldására hozott létre egy új eszközrendszert a meglévő módszerek újszerű kombinációjával. A dolgozatban a leíró jelleg kapott nagyobb szerepet, a módszerek hatékonyság elemzésére több kísérleti vizsgálatot is bemutat a szerző. Az értekezés alaposan és széleskörűen elemzi a kapcsolódó irodalmat.

A dolgozat szerkezete, logikai felépítése megfelelő. A leírásokban erős a módszertani megközelítés, a formális nyelvezeten viszont még maradt csiszolni való. A dolgozat angol nyelven készült, a jelölt magabiztosan uralja a szakmai nyelvet.

A szerző eredményeit megfelelően publikálta, több impakt faktoros cikk társszerzője.

Az értekezést elfogadásra javaslom.

Miskolc, 2021.12.04



Prof. Kovács László