

Bírálati vélemény a
"Network science-based analysis of human factors in systems engineering"
című doktori értekezésről

A digitalizáció elterjedésével és a rendelkezésre álló számítógépes kapacitásoknak köszönhetően igen részletes és lefedettségét illetően akár országos méretű adatbázisok készülhetnek. A dolgozat három különböző területen mutatja be, hogy a hálózat-tudományi módszertan miként alkalmazható a nagy tömegű adatban rejlő információ kiszűrésére és egy döntéshozó számára legfontosabb jellegzetességek kiemelésére. A dolgozat részletesen bemutat három elemzést, amelyekben a lakosság képzettségéről, munkavállalásáról, a gazdasági szereplők elhelyezkedéséről és vállalatok alkalmazottainak összetett viszonyairól mutat be egy-egy lehetséges elemzési példát. A felhasznált adatbázisok egyediak, és a hazai oktatás/munkavállalás/hatékonyság-növelés aktuális problémáinak elemzésére alkalmas modern eszközöket alkalmaz. Ezek alapján a választott témát és a bemutatott megoldásokat maximálisan időszerűnek találom.

A disszertáció összesen 3 tartalmi és egy bevezető fejezetből áll. A dolgozatot egy összefoglaló fejezet zárja, amely, a bevezetőhöz hasonlóan, megmutatja, hogy a látszólag különböző témákat mi kapcsolja össze egy egésszé, mi az az általános következtetés, amely éppen azért vonható le, mert a jelölt munkája egy széles spektrumot fed le alkalmazások szempontjából. Az 1. fejezet áttekintő bevezetése után a 2. fejezet bemutat egy csoportosítási módszertant, amellyel oktatási programokat és munkahelyi feladatokat lehet összekapcsolni. Ezt követően a 3. fejezet geográfiai elhelyezkedés szempontjából elemzi a hazai cégtulajdonosi hálózatot. A 4. fejezetben egy vállalat-szervezést felmérő kérdőív eredményeivel értékeli a munkahelyi vezetők és a munkavállalók együttműködésének minőségét és a rejtett vezetői szerepek felszínre hozásának lehetőségét.

A dolgozat különös értékének tartom, hogy egymástól távolinak tűnő területeket kapcsol össze (mérnöki rendszer-tervezés és humán erőforrás-kezelés) a hálózat tudomány eszköztárának alkalmazásával. Valószínűleg ez a megközelítés a kutatócsoportban természetesnek tűnhet, egy külső szemlélő számára azonban igen érdekes társítás, ami akár önálló diszciplínává is fejlődhet.

A jelölt a munkájának eredményeit három tézis kimondásával foglalja össze annak megfelelően, ahogy a dolgozatban három tartalmi fejezetre tagolta a munkásságát. Az egyes pontok megfogalmazását azonban nagyon általánosnak érzem. A vastaggal szedett tézispontok az adott téma feldolgozását lehetővé tévő módszertant nevezik meg. A tézispontok kifejtésében, az „Eredmények” pontban találtam olyan megfogalmazásokat, amely a jelölt munkáját és konkrét megoldásait, levont következtetéseit tartalmazzák. Feltételezem, hogy ez a helyi szokásoknak megfelelő igény miatt készült így. Az egyes pontok „Eredmények” bekezdésében megfogalmazott határozott kijelentések miatt a téziseket is elfogadhatónak tartom, a jelölt munkájának eredményeként ismerem el.

Véleményem szerint a dolgozat felépítése logikus, a témák kibontása és az eredmények ismertetése megfelelő és a szöveg megfelel az angol nyelv nyelvtani és helyesírási követelményeinek egyes gépelési és újraserkesztési hibáktól eltekintve. Bár a hibák észrevehetőek, a tartalmi megértést nem akadályozzák. A témaválasztásnak köszönhetően a dolgozatban sok leíró részlet található, de az alkalmazott elméleti háttér is alaposan és tömören ismertetésre került. A jelölt nagyon jó egyensúlyt talált az elméleti módszertan, az adatok és elemzések bemutatása és a témánkénti következtetések levonása között. Véleményem szerint az alkalmazott statisztikus- és hálózatelméleti módszerek korszerűek és magas színvonalúak. A dolgozat felépítésében és megfogalmazásában is törekszik a jól követhetőségre. Ezek alapján megállapítható, hogy a jelöltnek fontos hozzájárulása volt a választott tudományterület fejlődéséhez. Ezt maximálisan alátámasztja a jelölt társszerzőségével közölt több publikáció, melyek rangos konferencia kivonatokban, illetve referált tudományos folyóiratokban jelentek meg. Külön kiemelandó, hogy a dolgozat alapjául szolgáló, és a téziszüzetben megfogalmazott állítások mindegyike olyan tudományos publikációként jelent meg, melynek a jelölt az első szerzője. Bár mind a négy, nemzetközi, rangos folyóiratban megjelent publikációra igaz ez a megállapítás a

negyedik cikket nem hivatkozta a szerző a dolgozatban. Fontosnak tartom azt is megjegyezni, hogy a nemzetközi megjelenéseken túl jelentős hazai tevékenységet is végzett a jelölt.

Az értekezéssel kapcsolatos kérdéseim, megjegyzéseim a következők:

- Az oktatási programok és munkahelyek kapcsolásakor lehetségesnek tartja-e olyan mintázatok kibányászását, amellyel esetleg egy-egy személy beazonosítható lenne? Milyen módszert javasolna arra, hogy egy adatbázis anonimitását ellenőrizni lehessen?
- A 2.3 és 2.4 ábrán bemutatott kereset-különbséget meg tudja-e magyarázni, vagy legalább előzetes jelét látja-e a tanulmányok során befutott életpályákból vagy az adatbázis egyéb jellemzőiből?
- Hogyan változnak a 2. Fejezetben meghatározott klaszterek jellemzői, ahogy az alfa paraméter változtatásával ritkul a hálózat? Vannak-e olyan elemek, amelyek átvándorolnak egyik klaszterből a másikba? Ha igen, hogyan magyarázható ez a vándorlás?
- A 3.9 és 3.10 ábrákon egy-egy település külön ki van írva. Mi a szerepe ezeknek a településeknek, miért pont ezek vannak kiírva?
- A dolgozatban bemutatott eredmények mennyire robusztusak a kiválasztott csoportkeresési eljárási algoritmusra nézve? Például a 3.13-as ábra eredménye mennyire lenne reprodukálható más klaszterezési eljárásokkal?
- A 4-ik fejezetben bemutatott eredményeknél az egyes cégekre külön-külön készültek elemzések. Ez egy természetes hozzáállás. De ha az anonim adatokat összekeverjük a cégek között, lehetséges-e létrehozni egy „átlagos” céget, ahol az egyedi cégekben megállapított pozíciójú személyek vannak összevonva, és a kapcsolataik átlagolva? Elképzelhetőnek tartja-e, hogy ha sok cégről rendelkezésre állna ilyen jellegű kimutatás, akkor a cégek összeátlagolásával létrehozható lenne-e egy tapasztalati null-modell? Hogyan próbálna létrehozni egy ilyen, akár referenciaként használható, átlagos céget?

Az értekezéssel kapcsolatos apróbb, technikai jellegű kérdések, megjegyzések:

- Feltehetőleg a szövegalkotás folyamán megváltozott a szöveg tördelése, így néhány elválasztó-jel a sorok közepére került a tézis-füzetben. Esztétikai hátrányon kívül azonban nem okoznak ezek problémát, bár egy ilyen rövid, de mégis fontos részre nagyobb gondot lehetett volna fordítani.
- A disszertációban egyetlen zavaró megfogalmazást találtam: a 4-ik fejezetben a vásárlói kosár-formalizmusnál a jelölések azt indukálják, hogy A és B eseményeket jelöl pl. $P(A|B)$, azonban itt A és B dimenziókat jelöl. Pl a konfidencia definíciójánál az $A \cup B$ jelölés eseményekkel megfogalmazva A metszet B lenne. Bár ez egy tankönyvi szintű klasszikus probléma, de a dolgozat jellege miatt várhatóan olyanok is olvassák, akik nem jártasak az adatbányászat szabály-kinyerési zsargonjában, és ezért hibásnak gondolhatják a leírást.

Javaslom az értekezés nyilvános vitára bocsátását, a dolgozatot **summa cum laude** elismeréssel értékelem.

Előzetesen javasolom. PP

Budapest, 2020. október 30.



Pollner Péter