

Bírálat
Karsai Mártonnak az MTA Doktora címért benyújtott
„*Computational studies of human dynamics*”
című disszertációjáról

A dolgozat a komplex hálózatok valamint az emberi viselkedés dinamikájának vizsgálatáról és modellezéséről mutat be fontos eredményeket. A komplex hálózatok tanulmányozása az elmúlt két évtizedben egy nagyon népszerű és szerteágazó interdiszciplináris tudományterületté fejlődött, melynek fontosságát nem kell különösképpen hangsúlyozni. A nagyszámú, egymással kapcsolatban álló vagy egymással kölcsönható részegységből álló rendszerek fellelhetők szinte minden skálán, az egyes sejteken belül lejátszódó biokémiai reakcióktól kezdve a genetikai és fehérjekölcsönhatási hálózatokon, ideghálózatokon keresztül egészen az online szociális média platformokig vagy akár az egész emberi társadalomig. Ezen rendszerek tanulmányozhatóságát jelentősen segítette a digitális technológia ezredforduló táján elinduló rohamos fejlődése. Ennek köszönhetően manapság sokkal könnyebben lehet a különféle rendszerekről adatokhoz hozzáférni mint korábban, ugyanakkor a számítógépek teljesítménye is elérte azt a szintet, mely lehetővé teszi a hálózati analízist végző algoritmusok értelmes időkeretben történő lefuttatását. Ez nagyban hozzájárult ahhoz, hogy a hálózat kutatással foglalkozó új tudományos publikációk száma mára már vetekszik a fizikához vagy kémiához sorolható új cikkek mennyiségével. A dolgozat szerzőjének munkássága kiválóan illeszkedik ebbe a trendbe. Fontos azonban megjegyezni, hogy itt többről van szó, mint csupán egy prosperáló területhez történő egyszerű hozzájárulásról, hiszen több olyan szűkebb, ugyanakkor nagyon fontos kutatási irány is felbukkan a dolgozatban mely esetén a szerző hozzájárulása alapvető fontosságúnak bizonyult a vonatkozó kulcs publikációk tudományos fogadtatása alapján.

A dolgozat 5 fejezetre tagozódik, melyből az első az általános bevezetést, a terminológia valamint a tanulmányozott adatrendszerek ismertetését tartalmazza. Az érdemi eredmények 3 fejezetben kerülnek bemutatásra, melyek mindegyike külön bevezető alszekcióval rendelkezik.

A 2. fejezet az emberi viselkedés azon aspektusára koncentrál, melyet angol szóval „*bursty*”-ként szokás jellemezni és nincs jó magyar megfelelője. Lényege, hogy az egymást követő események mint például a telefonhívások időbelisége erősen és nem-triviális módon inhomogén, ahol a hosszabb eseménytelen időszakokat rövidebb de nagyon sűrű eseménybekövetkezéssel rendelkező időszakok válthatják. Az ilyen folyamatok kvantitatív jellemzése sem teljesen magától értetődő, a manapság használatos mennyiségek egy részét a dolgozat szerzője és munkatársai vezették be, továbbá rámutattak a hagyományos, pusztán a korrelációs függvény lecsengésére alapozó megközelítés hiányosságaira.

A 3. fejezet időfüggő hálózatokkal foglalkozik, melyekben a rendszer felépítése és az egyes elemek tulajdonságai időben változhatnak, hasonlóan az olyan, mindenki által megtapasztalható hálózatokhoz mint például az online szociális média platformok világa, a telefonhívások hálózata vagy akár egy hosszabb távon általánosan az emberi is-

merettségek, barátságok hálózata. Bár a növekedés, mint az egyik legegyszerűbb módja a változásnak már a Barabási-Albert modellben is kulcsszerepet játszik, a hálózatelmélet ezen egyik alapmodelljéhez képest a valós időfüggő komplex rendszereket megragadó hálózatok sokkal bonyolultabbak. A dolgozat 3. fejezetében bemutatott eredmények alapján nyugodtan kijelenthetjük, hogy a dolgozat szerzője ezen a területen is alapvető fontosságú hozzájárulást könyvelhetett el mind az elemzés módszertan, mind az elmélet és modellalkotás szempontjából. Statisztikus fizikai nézőpontból egy különösen szép eredmény annak megmutatása, hogy az időfüggő hálózatokban az óriáskomponens fázisátalakulás-szerű megjelenése az irányított perkoláció univerzalitási osztályába esik.

A 4. fejezet a komplex hálózatokon lezajló kollektív viselkedési folyamatokról szól, azon belül elsősorban a terjedési folyamatokra koncentrálva. Az ilyen jelenségek vizsgálatának relevanciája mindenki számára egyértelművé vált a Covid-19 világjárvány során, melynek modellezése és lehetséges kontrollja motivációs célként bukkan fel a fejezet alapjául szolgáló cikkek egy részénél. A dolgozat szerzője részt vett a „MASZK - Magyar Adatszolgáltató Kérdőív” fejlesztésében is, melynek célja a magyar lakosság járványterjedéssel kapcsolatos viselkedésének (pl. távolságtartás) monitorozása volt, önkéntes önbevallásos alapon. A fejezet a megszokott, már egy fertőző szomszéd által átadható fertőzési folyamatok mellett összetettebb terjedési folyamatokat is tárgyal, ahol például a fertőzött szomszédok számának el kell érni egy kritikus küszöböt a terjedéshez, ami meglehetősen összetett viselkedéshez vezethet. A korábbi fejezethez hasonlóan a 4. fejezet esetén is megállapítható, hogy a dolgozat szerzőjének több lényeges hozzájárulása is volt a vizsgált terület fejlődéséhez. A dolgozat végül az 5. fejezettel záródik, melyben a szerző jelenlegi és tervezett jövőbeli kutatási irányairól kaphatunk egy rövid összefoglalót.

A dolgozattal kapcsolatos érdemi kérdéseim és megjegyzéseim a következők:

1. A 2.4.1 fejezet a humán egocentrikus kommunikációs hálózatok „*bursty*” jellegével foglalkozik, ahol azt olvashatjuk, hogy a kommunikációs események sorozatában fellelhető „*bursty*” jelleg egyik intuitívnek tűnő magyarázata lehetne az, hogy egy egyén által kezdeményezett, de több különböző másik személlyel folytatott interakciók között erős időbeli korreláció léphet fel. Azonban az empirikus adatok elemzése erre rácsfol és azt mutatja, hogy a „*bursty*” jelleg elsősorban az egyes éleken zajló kommunikáció sajátosságainak következménye. Ezzel kapcsolatban szeretném kérdezni azt, hogy mennyire jó modellje az emberi kommunikációnak az, ha egy „ego” éleit egymástól függetlennek tekintjük? Van-e bármi olyan effektus amihez mégis szükség van az egy csúcspól induló de különböző éleken folytatott kommunikációk közti korrelációra?
2. Formálisan úgy tűnik, mintha gond lenne a (2.15) egyenlettel: Az egyenlet felett lévő szöveg nem mondja, hogy b_e függene E -től és a b_e -re megadott formula sem mutat explicit E függést. Azonban ha b_e nem függene E -től, akkor a (2.15) egyenletben kiemelhetnénk a szumma elé ami biztosan nem a kívánt eredményre vezetne. Esetleg itt a b_e (véltetően E -től függő) definícióját szerencsés lenne ismertetni.

3. a 2.10-es ábra f) paneljén az $E \rightarrow 0$ határesetnél a görbék látszólag $2/3$ -hoz tartanak, illetve a g) panelnél és a 2.12 ábra c) paneljénél meg $3/4$ -hez ugyanebben a határesetben. Van-e ezeknek a viselkedéseknek egy egyszerű szemléletes magyarázata? Illetve van-e olyan triviális kommunikációs mintázat, amikor a teljes E tartományon $2/3$ illetve $3/4$ -en marad a vonatkozó érték?
4. A 2.12 ábra c) paneljén úgy tűnik, mintha a kék négyzetből két görbe is lenne, egy a sárga négyzet görbáját szorosan követve, illetve egy másik görbe ami valamivel lejjebb halad. Ez két külön görbe, vagy pedig egy darab erősen fluktuáló görbe, ami hol felugrik a sárga négyzet görbéjéhez közel, hol vissza a lenti burkolóra? Akármelyik eset is áll fent, mi ennek a magyarázata?
5. A 3.3 alfejezet időfüggő hálózatokon zajló SI terjedési folyamatok érdekes tulajdonságait mutatja be. A statikus hálókon zajló terjedési folyamatokhoz képest itt egy fontos különbség, hogy a terjedésnek tiszteletben kell tartania a kapcsolatok megjelenésének időbeliségét is. Ezzel kapcsolatban kérdezném meg, hogy a vizsgált hálózatoknak van-e „időtükrozesi szimmetriája” a következő értelemben: Tegyük fel, hogy először megfordítjuk az adatokban az időbeliséget, ami olyan, mintha ha a hálózat időfejlődését visszafelé játszánánk le. Ezután ezen a megfordított hálón végezzük el ugyanazokat az SI terjedési modellre épülő vizsgálatokat mint a dolgozatban. Vajon az így tapasztalt viselkedés ugyanaz lesz (statisztikai értelemben) mint eredetileg, vagy jól elkülöníthetően más?
6. A 3. fejezet végén az egocentrikus háló időfejlődésével kapcsolatos eredményekről, modellekről olvashatunk. Ezzel kapcsolatban szeretném kérdezni azt, hogy az egocentrikus háló időbeli bővülését összekapcsolták-e már korábban a nyelvészetben ismert Heaps-törvénnyel (vagy Heaps–Herdan-törvénnyel), mely azt mondja ki, hogy az egymástól különböző szavak száma egy dokumentumban a dokumentum méretétől hatványszerűen függ? Első ránézésre egy olyan háló bővülést leíró modell melyben az ego véletlenszerűen vagy már meglévő kapcsolatokat erősít meg, vagy teljesen új kapcsolatokat hoz létre nagyon hasonlít azon egyszerű szöveggeneráló modellekre, melyben véletlenszerűen vagy egy már használt szóval folytatjuk a dokumentumot, vagy egy teljesen új szót írunk le.
7. A 4. fejezet érdekes eredményeket mutat be a magyarországi Covid-19 terjedés első két hullámának összehasonlításáról. Ennek kapcsán felmerül az az aspektus is, hogy esetleg a második hullám alatt általánosan lazább lehetett a fegyelem a lakosság részéről, hiszen az első hullámot már átvészelte az ország. Voltak-e erre utaló jelek bármelyik hozzáférhető adatforrásban (pl. MASZK), illetve hogyan befolyásolja az eredmények értelmezését ha egy ilyen lehetőséget is figyelembe akarunk venni?

Összességében a doktori értekezés magas színvonalú és nagyívű tudományos munkásságot tükröz. Véleményem szerint az elért eredmények messzemenően teljesítik azt a szintet ami az MTA Doktora cím odaítéléséhez szükséges, függetlenül a fenti kérdésekre,

megjegyzésekre adott válaszoktól. Fontosnak tartom megemlíteni, hogy Karsai Márton nemzetközileg egy rendkívül jól beágyazott kutató aki publikációinak döntő többségét külföldi társszerzőkkel közölte. Emellett a tudományos közéletben is igen aktív, legutóbb például a NetSci 2023 konferencia (ami legrangosabb évenként megrendezett nemzetközi hálózatkutató konferencia) egyik vezető szervezője volt. Továbbá saját konferencia-tapasztalataim alapján biztosan állíthatom, hogy Karsai Márton munkásságát a komplex hálózatokkal foglalkozó tudományos közösség széles körben ismeri és elismeri. A fentiek alapján a nyilvános vita kitűzését javaslom.

A dolgozattal kapcsolatosan további apróbb technikai észrevételeim is vannak, melyek nem befolyásolják a doktori mű értékelését, viszont amennyiben később a dolgozat szélesebb körben is elérhető lesz, érdemes lenne kijavítani őket. Ezeket függelékszerűen a bírálat végén sorolom fel.

Budapest, 2023. szept. 7.



Palla Gergely

A dolgozattal kapcsolatos apróbb, technikai megjegyzéseim:

- A tartalomjegyzékben nem jelenik meg a fejezetek címe.
- Az 1.1 fejezetben az 5. oldalon a „*Modelling*” alfejezet első mondata szerintem nyelvtanilag nem teljesen helyes (hiányzik belőle az állítmány).
- Elgépelés a 13. oldalon az „*Other datasets*” első mondatában, az „*I addition*” szerintem helyesen „*In addition*” lenne.
- Elgépelés a 23. oldalon a (2.13) egyenlet alatt 3 sorral: „*constant average event rate as shown in as demonstrated in Fig.2.6b.*”.
- Elgépelés A 49. oldalon a 3.7 ábránál az ábraalírásban: a második „(c)” helyesen „(d)”.
- Az 52. oldalon a 3.9 ábra a) paneljében a bal oldalon a sorozat utolsó motívumában hiányzik a 2-es a visszafelé irányuló nyílról.
- Az 55. oldalon a (3.7) egyenlet felett a szöveg átlagos méretet mond, de ez igazából a második momentum, nem az első.

- Elírások az 58. oldalon: „*During the our earlier analysis*”, valamint „*the order parameter was measured as the size (in terns if nodes and events)*”, továbbá „*Finally, by incorporated the effects of an external field h to this scenario as the spontaneous emergence of sources of infection modeled as an independent Poisson point process*”.
- A 3.5 fejezetben többször előfordul, hogy amit a szöveg (4.3)-as egyenletként hivatkozik (úgy mint „*Eq. 4.3*”) az valójában a (3.14)-es egyenlet.
- A 65. oldalon a 3.14. ábra d) paneljének felirata (MPC) és az ábraalírásban hivatkozott rövidítés (MPN) eltér (ez utóbbi valószínűleg elírás). Hasonlóan a h) panelen látható TWT rövidítés az ábraalírásban már TMN-ként szerepel.
- A 74. oldalon az első bekezdésben előfordul egy $p_{\text{rand}}(n)$ mennyiség ami valószínűleg helyesen $\pi_{\text{rand}}(n)$ lenne.
- A 85. oldalon a 4.5. ábra a „*transmission centrality*” mennyiséget szemlélteti. Ennek kapcsán a szövegben vázolt alapértelmezés szerint az SI modell $\beta = 1$ -es esetét kell használnunk a centralitás kiszámításához. Egy ilyen folyamat esetén (a $\beta = 1$ miatt) azonban a fertőzés kiinduló gócaból mindenkihez a legrövidebb úton jut el a fertőzés, szemben azzal amit a 4.5. ábra mutat, ahol sok csúcsba egy hosszabb úton jutott el a fertőzés. Itt mindenképp jelezni kellene, hogy az ábra egy általánosabb esetet mutat ahol a $\beta < 1$.