

Budapesti Műszaki és Gazdaságtudományi Egyetem
Villamosmérnöki és Informatikai Kar
Távközlési és Médiainformatikai Tanszék
High-Speed Networks Laboratory (*HSN Lab*)
MTA-BME Információs rendszerek kutatócsoport

Funkció és Struktúra Összefüggése Komplex Hálózatokban

MTA doktori disszertáció tézisei

Gulyás András, Ph.D.

Budapest
2020

1. Bevezető

A valós (komplex) hálózatok kutatása egyike a leginkább multidiszciplináris tudományterületeknek. Abból indul ki, hogy szinte minden valós rendszer (Internet, közösségi háló, szótárak, sejt, közlekedés, emberek közötti interakciók, Linux csomagok, Java programkönyvtárak stb.) képzelhető egy hálózatnak, gráfnak és pusztán a hálózat struktúrájának ismerete alapján sok mindent megállapíthatunk ezen rendszerek viselkedéséről. Ráadásul, az ezredforduló egyik érdekes megfigyelése volt, hogy a különféle rendszerekből származtatott hálózatok struktúrája sok tekintetben hasonló [1]. Ez azt jelenti, hogy pl. az Internet struktúrájából levont következtetések alkalmazhatók lehetnek más hálózatokban, pl. biológiai hálózatokban vagy gyógyszerkutatásban és viszont, az online közösségi rendszerek viselkedésének tanulmányozása inspirálhat valós számítógép-hálózati megoldásokat.

2. Kutatási előzmények és célkitűzések

A valós hálózatokkal kapcsolatos kutatások általában a hálózat struktúrájára és az abból adódó működési sajátosságokra koncentrálnak. A kétezres évek óta hatalmas mennyiségű mérési adat és tudás halmozódott fel a hálózatok szerkezetére vonatkozóan. Kiderült, hogy a legtöbb valós hálózat átmérője meglepően kicsi, foksámeloszlása hatványfüggvény alakú és a csoportképződési együtthatója sokkal magasabb, mint a véletlen hálózatoknak [2]. Ezeken a klasszikusnak számító strukturális tulajdonságokon túl a valós hálózatok sok egyéb strukturális jellegzetességeit feltárták. Centralitás eloszlások, asszortativitás, adjacencia-mátrix tulajdonságok csak néhány kiragadott példa olyan tulajdonság csoportokra amelyeket a kutatóközösség részletesen feltérképezett az utóbbi két évtizedben. A struktúra széles körű feltárása mellett rengeteg hálózatmodell született, amely valamilyen módszerrel képes a valós hálózatok szerkezetét reprodukálni [3]. A modellek túlnyomó többsége ún. generatív modell, ami azt jelenti, hogy a modell valójában egy szisztematikus hálózatgeneráló algoritmus, ami a valós hálózatokhoz adott tekintetben hasonló gráf-struktúrát állít elő.

A szerkezeti jellegzetességekből kiindulva meg lehet becsülni a hálózatok viselkedését különféle helyzetekben [3]. A struktúra ismeretében például, tudhatjuk, mennyire ellenálló a hálózat véletlen hibák, vagy szándékos támadás esetén, modellezhető, hogyan terjednek rajta a különféle vírusok, megvizsgálhatjuk, milyen módszerekkel lehet tájékozódni a hálózatban, de pl. azt is, hogy mely csomópontokra kell hatást

gyakorolnunk ahhoz, hogy a hálózat egészének a viselkedését megváltoztassuk. Ezek tulajdonképpen a hálózat struktúrájából következő hálózati funkciók, vagyis a szakirodalomban fellelhető legtöbb tanulmány a struktúra→funkció logikát követi: a hálózat szerkezetéből indulunk ki és meghatározzuk az abból adódó funkciókat.

Dolgozatomban megfordítom az érvelés menetét. Abból a hipotézisből indulok ki, hogy a hálózatok valamiféle funkcionális céllal jönnek létre és ebből a funkcióból kell, hogy következzen a hálózat szerkezete. Kiindulópontom tehát mindig valamilyen alapvető hálózati funkció, amely köré olyan modellt építek, amelyben a funkció képes hálózati struktúrát létrehozni. A modell tehát végül is hálózatot eredményez, melyet összehasonlítok a valós hálózatokkal.

A funkció→struktúra megközelítés véleményem szerint sokkal természetesebb megközelítése a hálózatok világának. A generikus modellek ugyanis nagyon jól megválaszolják a „hogyan” kérdését, vagyis azt, hogy miképpen tudunk a valós hálózatokhoz hasonló szintetikus hálózatokat létrehozni. Arra viszont nem képesek választ adni, hogy „miért” jönnek létre ilyen típusú hálózatok. A funkció→struktúra megközelítés pontosan ezt a „miért”-et feszegeti és a választ a funkcióval adja meg. Leegyszerűsítve: a hálózatok szerkezete azért olyan amilyen, hogy egy vagy több konkrét funkciót hatékonyan ki tudjon szolgálni. Dolgozatomban a hálózatok egyik legfőbb funkcióját az útvonalak kiválasztását (navigáció, keresés, információ terjedés, útvonalválasztás, routing) vizsgálom és a hálózati szerkezetre gyakorolt hatásait részletesen elemzem.

3. Kutatási módszerek

A hálózatokkal kapcsolatos kutatások nyelve természetesen a gráfelmélet, melyet széles körben használtam kutatásaimban és a dolgozatban is. A funkció→struktúra megközelítés fontos kérdése, hogy miért jönnek létre adott szerkezetű hálózatok, vagyis mi ösztönzi arra a csomópontokat, hogy egy adott struktúrát hozzanak létre? Egy adott populáció alapvetően önző elemeinek interakciójából kialakult viselkedést a játékelmélet írja le, ennek megfelelően a funkció→struktúra megközelítésben használt modelljeim a gráfok és a játékelmélet kombinációi. A modellek kiértékelését, ahol csak lehetséges analitikus úton végeztem, viszont néhány modell komplexitása csak a szimulációs elemzést tette lehetővé. Ezen felül nagyon érdekelt, hogy a valóságban miféle útválasztást használnak a különféle hálózatok az információ átvitelére, ezért méréseket, adatgyűjtést és adatelemzést is végeztem.

4. Új tudományos eredmények

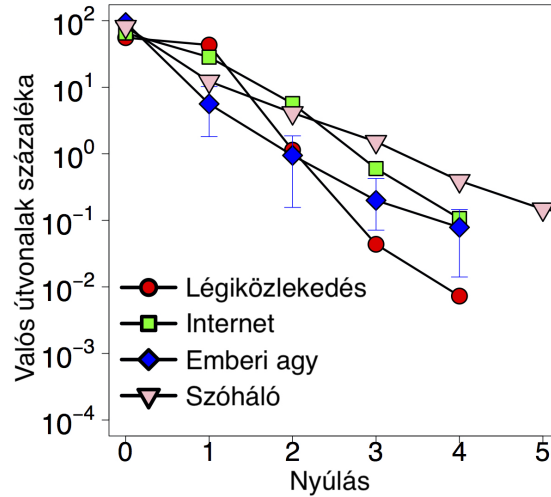
A dolgozatomban részletezett új tudományos eredmények három csoportba oszthatók. Az tézisek első csoportja magának a funkció→struktúra megközelítésnek a megalapozása és definíciója. A második és a harmadik csoport pedig két konkrét hálózati útválasztási módszer, mint funkció és az azokból következő hálózati szerkezet analízisét tartalmazza.

1. A funkció→struktúra megközelítés megalapozása

A hálózatok talán legnyilvánvalóbb célja, hogy a csomópontjai között lehetővé tegye a kommunikációt, [4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19]. Jelenleg a tudományos közösségben leginkább elterjedt feltevés a kommunikáció módjára vonatkozóan a legrövidebb út feltevés [20]. Ez azt jelenti, hogy két csomópont a közöttük levő összes lehetséges út közül a legrövidebben keresztül kommunikál. Ez a feltevés azért rendkívül népszerű, mert egyszerű és könnyen számolható, valamint sok esetben egyezik a tapasztalattal is. Nagyon kevés azonban a szakirodalomban fellelhető mérés [21, 22, 23, 24], vagy kísérlet amely a hálózatokban végbemenő kommunikáció valós útvonalait célozta volna. Márpedig az útvonalak ismerete nagyon fontos, hiszen ezek által ismerhetők meg a valós hálózatban alkalmazott útválasztási módszerek, amelyek a funkció→struktúra megközelítés kiindulópontját is adják. Méréseket és kísérleteket végeztem ezért négy egymástól nagyon különböző területről vett hálózaton az útvonalak természetére vonatkozóan. Több kutatóintézetrel együttműködve sikerült használható útvonal adatbázist összeállítanunk az Internetre, a légi közlekedési hálózatra, az emberi agyhálózatra, illetve egy szóhálózatra vonatkozóan. Az adatbázisok elemzése meglepő következtetésre vezetett.

1.1. Tézis ([25, 26]). *Az élet különböző területeiről (Internet, biológia, légi közlekedés, szóhálózat) vett hálózatokra vonatkozó mérések elemzésével megmutattam, hogy a valós hálózatokban az útvonalak számítása, a tématerületen uralkodó feltevéssel ellentétben, nem a legrövidebb útvonalválasztás algoritmusa szerint történik (1. ábra).*

Fontos kiemelni, hogy a 1.1. Tézis nem azt állítja, hogy sohasem megy a legrövidebb utakon a kommunikáció. A tézis állítása, hogy, egy olyan séma dolgozik a háttérben, ami gyakran adja vissza a legrövidebb utat, de semmiképpen sem minden esetben, sőt előfordulnak nagyon karakterisztikus eltérések a legrövidebb utaktól. Úgy tetszik tehát, hogy a hálózatok valamiféle, a legrövidebb út választásától különböző séma



1. ábra. Valós hálózati útvonalak nyúlása a legrövidebb úthoz képest. Nyúlás alatt itt a valós útvonal (élszámban vett) hosszának és a legrövidebb út hosszának különbségét értjük. Bár a valós útvonalak többsége esetében a nyúlás 0, az útvonalak 20-40%-a nyúlik. A nyúlás sok esetben eléri a 4-5-öt is. Az ábráról leolvasható, a nyúlás eloszlás meglepő hasonlósága a négy nagyon különböző hálózatban, ami felveti annak lehetőségét, hogy a különféle hálózatok hasonló útválasztási sémákat használhatnak.

szerint bonyolítják le a kommunikációt. Ahhoz tehát, hogy a lehetséges útválasztási módszerek szerkezetre gyakorolt hatását elemezni tudjuk egy olyan játékelméleti modell felállítására van szükség, ami rugalmasan tudja kezelni a lehetséges kommunikációs sémákat és meg tudja mutatni azok hálózati szerkezetre gyakorolt hatását.

A dolgozatomban definiált játékelméleti modell a csomópontok (játékosok) egy halmazának ($u = 1, 2, \dots, N$) egymásrahatásából következő egyensúlyi állapotként írja le a hálózatot. Egy u csomópont lehetséges stratégiáinak tetszőleges számú él behúzása u -ból indulva tetszőleges más csomópontokba. Egy szélső helyzet például, hogy u egyetlen élet sem húz be, mint ahogy az is, hogy minden lehetséges csomópontba csatlakozik. A lehetséges stratégiák száma egy csomópontra 2^{N-1} . Ezen stratégiák kombinációjából egy hálózat származtatható N csomópont felett. Egy u csomópont célja, hogy a lehető legkevesebb él behúzásával megoldja, hogy a hálózat tetszőleges v csomópontja felé navigációval el tudjon jutni. Formálisan, u költségfüggvénye, melyet minimalizálni igyekszik: $c_u = k_u + n_u$, ahol k_u az u által alkalmazott élbehúzás költsége, n_u pedig megadja a kommunikáció költségét az összes többi csomópontba. A modell egzakt megfogalmazását az 1.2. Tézis adja

meg.

1.2. Tézis ([27, 28, 29, 30, 31]). *Megadtam egy játékelméleti modellt, amely a kommunikációs költség alkalmas megválasztásával rugalmasan képes kezelni a lehetséges kommunikációs sémákat és fel tudja fedni azok hálózati szerkezetre gyakorolt hatását, a következők szerint. Egy $u \in \mathcal{P}$ csomópont lehetséges stratégiái az élek egy halmaza: $S_u = 2^{\mathcal{P} \setminus \{u\}}$. Az összes csomópont stratégiáját alkotó s vektor $s = (s_0, s_1 \dots s_{N-1}) \in (S_0, S_1 \dots S_{N-1})$ egyértelműen megadja a $G(s)$ gráfot: $G(s) = \bigcup_{i=0}^{N-1} (i \times s_i)$. Egy u csomópont költségfüggvénye:*

$$c_u = \sum_{\forall u \neq v} d_{G(s)}(u, v) + k(s_u), \quad u, v \in \mathcal{P} \quad (1)$$

ahol $d_{G(s)}(u, v)$ a kommunikáció költsége u -ból v -be a $G(s)$ gráfon, $k(s_u)$ pedig az s_u stratégia megvalósításának költsége.

2. Navigálható hálózatok funkció→struktúra analízise

A hálózatok navigálhatóságának fogalmát a szakirodalom egy adott geometriai sík felett működő mohó keresési algoritmus sikerességével azonosítja. Az algoritmus mohósága abban rejlik, hogy minden csomópont lokális döntést hoz arról, hogy milyen irányba küldje tovább az információt. Tehát egy u csomópont úgy továbbítja egy v -nek szóló információt, hogy a szomszédai közül kiválasztja azt a v' -t, amelyik a legközelebb van v -hez (esetleg pont v -t) az adott geometriai síkon. Ha u -nak nincs olyan szomszédja, amely közelebb van v felé, mint u saját maga, akkor azt mondjuk, hogy u nem képes v -hez navigálni. Azon u, v csomópontok százalékos aránya melyek között lehetséges mohó navigációval közlekedni adja meg a ún. sikerességi arányt. A mohó navigációhoz igazodva pontosítottam tehát a funkció→struktúra játékelméleti modellt. Az élbehúzás költségét egyszerűen az élek számával azonosítom ($k(s_u) = |s_u|$), a navigáció költségét pedig a következőképpen adom meg:

$$d_{G(s)}(u, v) = \begin{cases} 0 & \exists u \rightarrow v \text{ mohó útvonal a } G(s) \text{ gráfon} \\ \infty & \text{egyébként.} \end{cases}$$

A modell ebben a formában elég általános és bármilyen geometriát képes kezelni. A valós hálózatok rejtett geometriája azonban legtöbb

esetben nem euklideszi, hanem Bolyai-féle hiperbolikus geometria [32]. N csomópontból ($u = 1, 2, \dots, N$) indulok tehát ki, melyeket véletlenszerűen szórunk le egy R sugarú hiperbolikus körlap felületére. A pontok polárkoordinátáihoz (r, ϕ) , $r \in [0, R]$, $\phi \in [0, 2\pi]$, tartozó sűrűségfüggvények [33]:

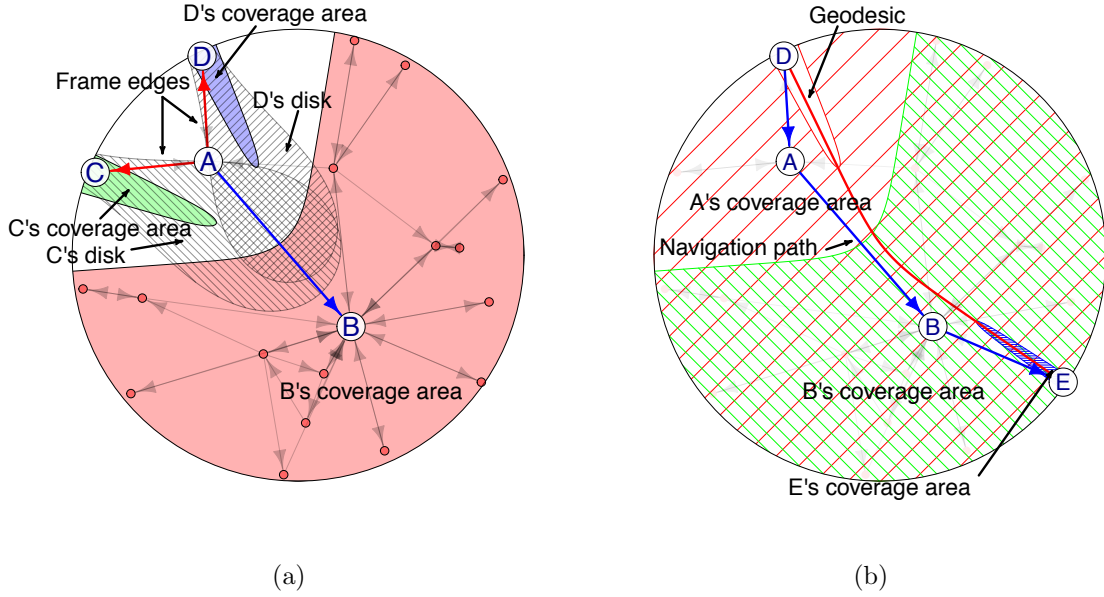
$$\rho(r) = \frac{\alpha \sinh(\alpha r)}{\cosh(\alpha R) - 1}, \quad \rho(\phi) = \frac{1}{2\pi}, \quad (2)$$

ahol $\alpha > 1/2$ a pont-leszórás heterogenitását szabályozza. A hiperbolikus távolságot a sík tetszőleges u és v pontjai között a következő formulával számoljuk:

$$d(u, v) = \operatorname{acosh} [\cosh r_u \cosh r_v - \sinh r_u \sinh r_v \cos(\phi_u - \phi_v)]. \quad (3)$$

Ezzel a geometriai kiegészítéssel definiált játékelméleti modellt hálózati navigációs játéknak (NNG) nevezem, melynek illusztrációja a 2. ábrán található. Az (a) panel az A csomópont optimális stratégiáját (B, C, D) mutatja egy kisméretű ponthalmaz esetén. Optimális, mert ennél kevesebb éllel nem biztosítható a navigáció a hálózat összes többi pontjához, ami végtelen költséget adna A számára. A B, C és D pontok A számára a hálózat egészét navigációs szempontból lefedik, vagyis rajtuk keresztül az A pontból tetszőleges más pontba indulva a mohó algoritmus szerint tovább lehet lépni. Az A -ra vonatkozó ún. fedési területeket a színezett tartományok mutatják. B fedési területe A számára a pirossal színezett rész. Ezek a pontok a hiperbolikus körlapon közelebb vannak B -hez mint A -hoz, így bármelyikhez is szeretne A navigációval eljutni, a B pont választása mindig megfelelő. Ennek megfelelően a modell optimális megoldása egy adott csomópontban a minimális halmazfedési problémára vezethető vissza. Az AC (és AD) él ún. váz él, mert A fekszik legközelebb C -hez, mint ahogy ezt a C -köré rajzolt $|AC|$ sugarú hiperbolikus körlap is szemlélteti (rácsozott rész), ami csak C -t és A -t tartalmazza. Ezért, ha A , C -be akar navigálni, nincs más választása, mint közvetlen élet létrehozni C -hez. A (b) panel illusztrálja a csökkenő méretű fedési területeket a navigáció útvonalán D -ből E -be (kék nyilak). A piros görbe a hiperbolikus geodéta D és E között. A fedési területeket a navigáció irányában egyre növekvő sűrűségű vonalkázás jelzi.

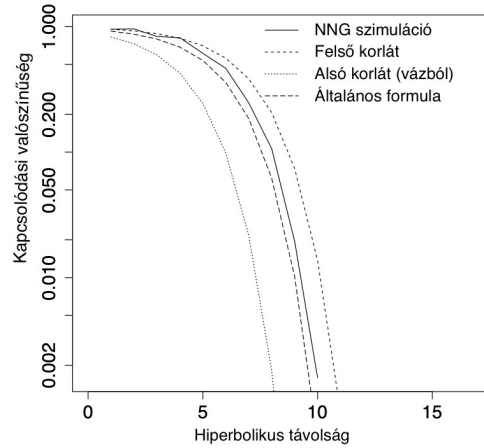
A navigációs játék jól egybefűzi a navigációt mint hálózati funkciót és a hálózat önszerveződését melyet a csomópontok költségfüggvényei irányítanak. Probléma viszont, hogy a modell következményeinek egzakt kiszámítása nehéz. A halmazfedési probléma NP-teljes [34] a játék lokálisan és globálisan is optimális egyensúlyi állapotának leszámolása



2. ábra. A navigációra vonatkozó játékelméleti modell (NNG) illusztrációja.

pedig N darab halmazfedési probléma megoldását feltételezi. Az egzakt megoldás kiszámolása ezért analitikusan nem megoldható, numerikusan pedig néhány ezer csomópontig számolható a jelenlegi szoftveres megoldókkal. Ha lemondunk az egzakt megoldásról és megelégszünk az egyensúlyi állapotról kapott magas szintű statisztikákkal és hálózati jellemzőkkel, átfogalmazhatjuk a problémát a statisztikus mechanika segítségével. Az egzakt megoldás helyett gráfsokaságokat vizsgálva lehetőségünk van statisztikák felállítására az átlagos foksám, a foksámeloszlás és a klaszterezettségi együttható tekintetében. A módszer nem teszi lehetővé az egyes élek megjelenésének pontos leírását, viszont a gráfsokaság felett értelmezett kapcsolódási valószínűség (p_{ij}) analitikus úton számolható. A kapcsolódási valószínűségek ismeretében pedig a hálózat globális jellemzői származtathatók.

2.1. Tézis ([27]). *Analitikus úton megmutattam, hogy a navigációs játék (NNG) mindig magában foglal egy navigációs vázat (greedy frame), amely minden lehetséges egyensúlyi állapotban megtalálható. Analitikus módszerekkel levezettem a vázhoz tartozó kapcsolódási valószínűségeket az összes lehetséges csomópontpárra, amely egyben alsó korlát is az egyensúlyi állapot kapcsolódási valószínűségére. Analitikus úton felső korlátot adtam az egyensúlyi állapot kapcsolódási valószínűségeire. Az alsó és a felső korlát felhasználásával általános formulát adtam a csomópontok közötti kapcsolódási valószínűségekre egyensúlyi helyzetben (3.*



3. ábra. Kapcsolódási valószínűség számítása a hiperbolikus távolság függvényében. Az ábrán látható az analitikus alsó és felső korlát illetve az általános formula (2.1. Tézis), valamint a szimulációból számított empirikus eloszlás is.

ábra).

A kapcsolódási valószínűségekre adott általános formula felhasználásával meghatároztam a navigációs játék egyensúlyi állapotához tartozó hálózat főbb jellemzőit.

2.2. Tézis ([27]). *Analitikus úton megmutattam, hogy navigációs játék (NNG) egyensúlyi állapotához ritka gráf tartozik (átlagos fokszám < 4), mely jó egyezést mutat a valós hálózatokkal. Szintén analitikus módon igazoltam, hogy az egyensúlyi hálózat fokszámeloszlása egyenletes pont-leszórás esetén hatványfüggvény ($\gamma = 3$). Analitikus módszerrel beláttam, hogy nem egyenletes csomópont-leszórással a hatványfüggvény kitevője változtatható. Szimulációs vizsgálatokkal igazoltam, hogy ($\gamma \approx 2$) esetben kapjuk a legalacsonyabb költséget miközben a hálózat maximálisan navigálható. Analitikus közelítéssel megmutattam, hogy az egyensúlyi hálózat klaszterezési együtthatója magas ($\bar{c} \approx 0.45$).*

2.3. Tézis ([27]). *Geometriába ágyazható valós (légiközlekedési hálózat, emberi agyhálózat, szóhálózat, Internet) hálózatok vizsgálatával igazoltam, hogy a navigációs játék (NNG) navigációs vázában, illetve egyensúlyi állapotában megjelenő élek nagy része ténylegesen (hálózattól függően 70-90%) megtalálható a valós hálózatban (1. táblázat). Szimulációs vizsgálatokkal beláttam, hogy a modell segítségével hatékonyan azonosíthatók azok a kritikus élek, melyek hozzáadásával illetve törlésével a valós hálózatok navigálhatósága nagy mértékben javítható, illetve rontható.*

	Inter. H	Metab. H	Szó H	Utak E	Légik. S	Légik. H	Agy E
Pontok száma	4919	602	4065	3136	283	283	998
Valós élek száma ($ R $)	28361	2498	38631	-	1973	1973	17865
NNG élek ($ M $)	5490	743	4634	9808	643	328	2591
Valódi pozitívak ($ T $)	4556	643	3311	8776	65	277	2306
Hibás pozitívak ($ F $)	934	100	1323	1032	578	51	285
Pontosság ($ T / M $)	83%	87%	71.5%	89.48%	10.1%	84%	89%
Váz élek ($ M_F $)	3680	415	3304	3105	199	249	716
Váz valódi pozitívak ($ T_F $)	3243	378	2528	2931	15	216	677
Váz pontosság ($ T_F / M_F $)	88%	91%	77%	94.40%	7.5%	87%	94.6%
Navigációs sikeresség	87%	85%	81%	-	54%	89%	89%

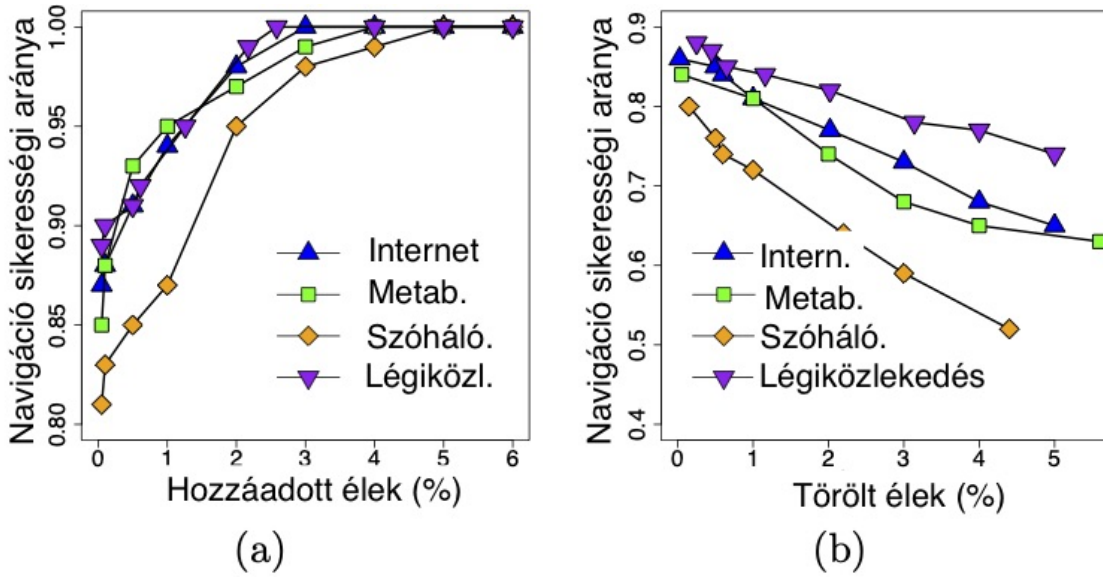
1. táblázat. Valós hálózatok és a hozzájuk tartozó NNG egyensúlyi hálózat éleinek összevetése.

Az 1. táblázatban látható az NNG által létrehozott élek és a valós hálózatokban található élek összevetése. A táblázat minden vizsgált hálózatra megadja a valós hálózatban található élek számát $|R|$, a hozzájuk tartozó NNG egyensúlyi hálózat élszámát $|M|$, a valódi pozitív élek számát $|T| = |M \cap R|$, a hibás pozitív élek számát $|F| = |M \setminus R|$, valamint a valódi pozitívak arányát $|T|/|M|$, melyet pontosságnak nevezünk. A táblázat a váz élek pontosság statisztikáit is tartalmazza, valamint a legelső sorban a navigáció sikerességét az egyes valós hálózatokban. A H,E,S nagybetűk a hálózatok nevei után a hálózatokhoz tartozó geometriát adja meg: H:hiperbolikus, E:euklideszi, S:gömbi.

A 4. ábrán látható, hogy az NNG egyensúlyi hálózat ismeretében hogyan javíthatunk, illetve ronthatunk a valós hálózatok navigálhatóságán. A 100%-ban navigálható NNG éleit a centralitásuk szerint csökkenő sorrendbe tesszük és sorrendben hozzáadjuk őket a valós hálózathoz, ha még nem szerepeltek ((a) panel), illetve kitöröljük, ha igen ((b) panel). Az x tengely a hozzáadott illetve törölt élek százalékos arányát mutatja a valós hálózat élszámához képest.

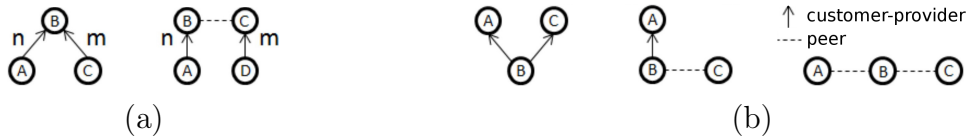
3. Navigáció és hierarchia

A hierarchikus hálózatok eklatáns példája az Internet. Az Internet esetében ugyanis az internetszolgáltatók közötti üzleti alapokon létrehozott összeköttetések két csoportba sorolhatók [35]. A *vásárló-szolgáltató* (customer-provider) kapcsolat esetén a vásárló (customer) szolgáltató fizet a tranzit szolgáltatónak (provider), hogy továbbítsa forgalmát, míg *egyenrangú* (peer) kapcsolat esetén a szomszédos internetszolgáltatók önkéntes alapon és ingyenesen cserélnek egymás között forgalmat. Ennek megfelelően az Interneten használt útvonalak sem lehetnek akármilyenek: a szolgáltatói hierarchiát be kell tartani! Ezt valósítja meg az ún. völgymentes szabály, ami azt mondja ki, hogy az A szolgáltató egy szomszédos B szolgáltató felé menő összeköttetését akkor és csak ak-



4. ábra. Valós hálózatok navigálhatóságának javítása, illetve lerontása az NNG felhasználásával

kor veheti igénybe, ha a bejövő forgalom A vásárlójától van vagy pedig B vásárlója A -nak. Kicsit formálisabban: a völgymentes útvonal állhat bármennyi (nulla is) vásárló-szolgáltató élből, utána legfeljebb egy egyenrangú élből, és újra bármennyi szolgáltató-vásárló élből, szigorúan ebben a sorrendben (Fig. 5).



5. ábra. Völgymentes (a) és nem völgymentes (b) útvonalak illusztrációja.

A völgymentességen kívül a legerterjedtebb, gazdaságilag is kívánatos útvonaltulajdonság a lokális preferencia. Ez azt jelent, hogy a szolgáltató az elérhető völgymentes útvonalak közül azt részesíti előnyben, ami egyenrangú, vagy szolgáltató-vásárló élel kezdődik, hiszen ezért az útvonalért nem fizet pénzt. Ha nincs ilyen útvonal, akkor választ csak vásárló-szolgáltató élel kezdődő utat, hiszen ezért fizetnie kell. A lokális preferencia a szolgáltatók elemi pénzügyi érdeke, ezért szinte minden szolgáltató alkalmazza.

Annak érdekében, hogy meg tudjam vizsgálni a völgymentesség és a lokális preferencia hatását az Internet szerkezetére vonatkozóan, pontosítottam a játékelméleti modellt a következők szerint. A költségfüggvény

egy u csomópont esetére:

$$C_u(s) = \underbrace{\frac{1}{N} \sum_{\forall v \neq u} d_{G(s)}(u, v)}_{\text{kommunikáció költsége}} + \underbrace{\varphi_p u_p + \varphi_r u_r}_{\text{élstratégia költsége}}, \quad v \in \mathcal{P} \quad (4)$$

ahol

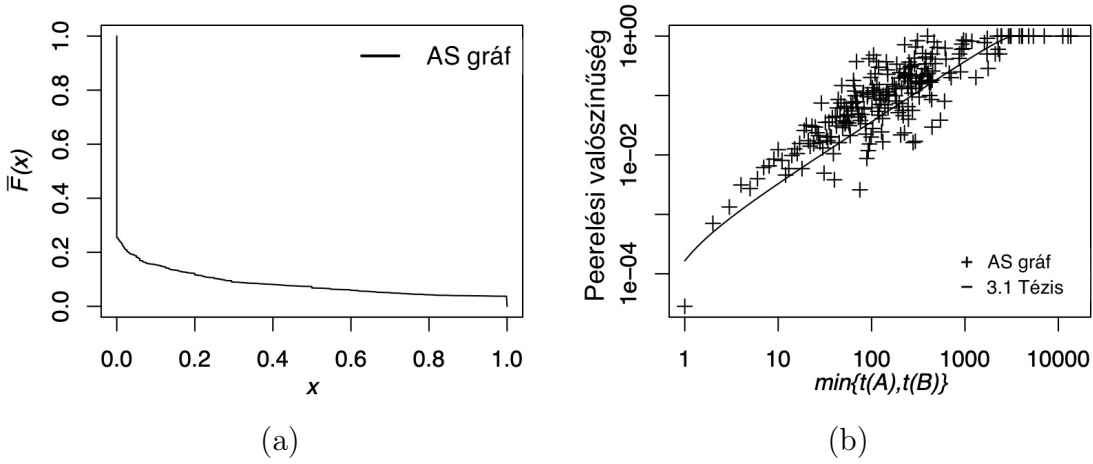
$$d_{G(s)}(u, v) = \begin{cases} 0 & \text{ha van völgymentes út, melyen az első él egyen-} \\ & \text{rangó or szolgáltató-vásárló} \\ 1 & \text{ha van legalább egy völgymentes út, melyeken az} \\ & \text{első él vásárló-szolgáltató} \\ \infty & \text{nincs völgymentes út} \end{cases} \quad (5)$$

adja meg a kommunikáció költségét u és v között $G(s)$ -felett, φ_p és φ_r a vásárló-szolgáltató és egyenrangú élek kialakítási költségei, valamint u_p és u_r adja meg a vásárló-szolgáltató és egyenrangú élek számát u -ban. Ezzel a hierarchikus kiegészítéssel definiált játékelméleti modellt hierarchikus hálózati játéknak (HNG) nevezem.

1. Definíció (Pókháló). Egy G gráfot pókhálónak nevezünk, ha kizárólag a következőket tartalmazza:

1. egy csupa egyenrangú élekből álló klikk K_r
2. olyan fák, amik gyökere $V(K_r)$ -ban van és csak vásárló-szolgáltató és egyenrangú éleket tartalmaznak, úgy, hogy a tranzit szolgáltató mindig közelebb van a gyökérhez mint a vásárló
3. plusz egyenrangú élek melyekre fennáll, hogy $\forall \bar{u}\bar{v}, \bar{u}\bar{w} \in G(s) : t(v) \cap t(w) = \emptyset$, ahol $t(x)$ az x csomópont alatti részében levő csomópontok halmaza x -et is beleértve.

3.1. Tézis ([31, 36]). Analitikus úton megmutattam, hogy a hierarchikus hálózati játékban (HNG) kialakult egyensúlyi hálózat mindig tartalmaz részgráfként egy pókhálót, amely kizárólag a völgymentes és lokális preferenciát használó útválasztás strukturális következménye. Analitikus módszerekkel igazoltam továbbá, hogy a modell jól megjósolja, hogy az egyenrangú élek körülbelül azonos méretű felhasználói táborral rendelkező szolgáltatók között alakulnak ki. A modell következtetéseit valós Internet-mérésekkel is igazoltam (6. ábra).

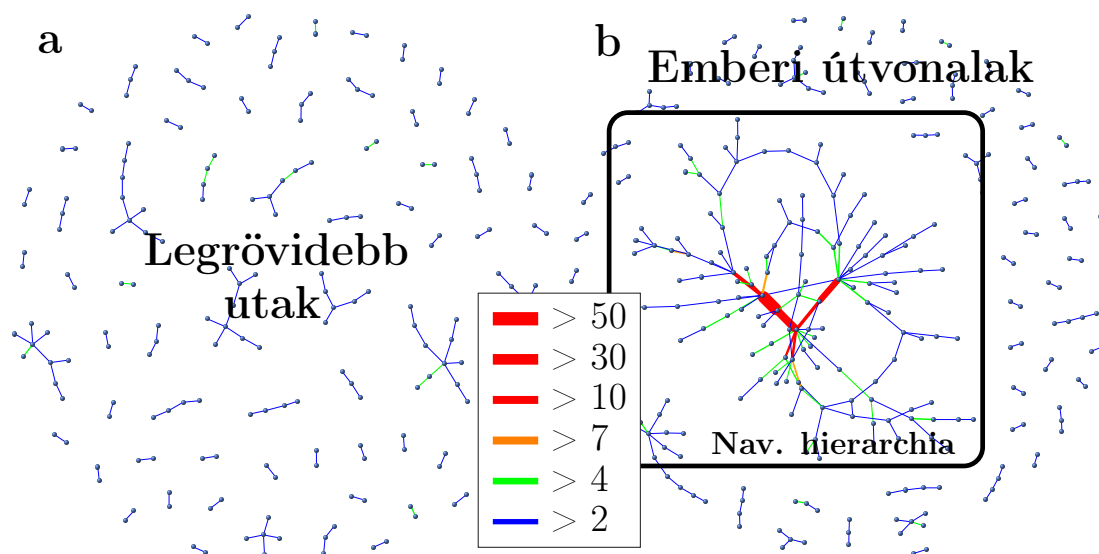


6. ábra. A peer élék végén levő internet szolgáltatók vásárlói között átfedés $x = \frac{t(A) \cap t(B)}{\min\{t(A), t(B)\}}$ kumulatív eloszlásfüggvénye ((a) panel). Peer él kialakulásának elméleti és tapasztalati valószínűsége adott vásárlói populációval rendelkező szolgáltatók között ((b) panel).

A hierarchikus hálózati felépítés nem csak az Internet esetében jellemző. Nagyon érdekes, hogy komplex hálózati rendszerekben az emberi tájékozódás szintén hierarchikus szerkezetekre épít.

3.2. Tézis ([25, 26, 37, 38, 39]). *Emberi alanyoktól felvett idősorokkal igazoltam, hogy az emberek esetében is hierarchia áll a navigáció hátterében. Megmutattam, hogy az emberek nagy mértékben leegyszerűsítik a hálózati tájékozódás folyamatát azzal, hogy teljes hálózat helyett annak egy leegyszerűsített (fához közeli) hierarchikus részgráfján oldják meg a navigációt (7. ábra).*

Az emberi útvonalválasztás és a legrövidebb út választásának különbségét illusztrálja a 7. ábra. Az (a) panelen látható, hogy a hálózat egyes éleit hányszor használta a legrövidebb utat választó algoritmus 1000 véletlenszerűen választott forrás-cél pár esetén. Megfigyelhető, hogy az élek használata meglehetősen homogén, hiszen a legrövidebb utat választó algoritmusnak nincs semmiféle belső interpretációja a hálózat struktúráját illetően. A hálózat szerkezetének ismeretében mechanikus módon kiválasztja a legrövidebb utakat. Az emberek viselkedése ettől szignifikánsan eltér. A (b) panelen emberi alanytól származó megoldás élhasználati statisztikája látható, ugyanarra az 1000 véletlen forrás-cél párra. Az emberi megoldások jól kivehető szerkezet köré csoportosulnak, ami annak a jele, hogy az embernek jól megfigyelhető koncepciója alakul

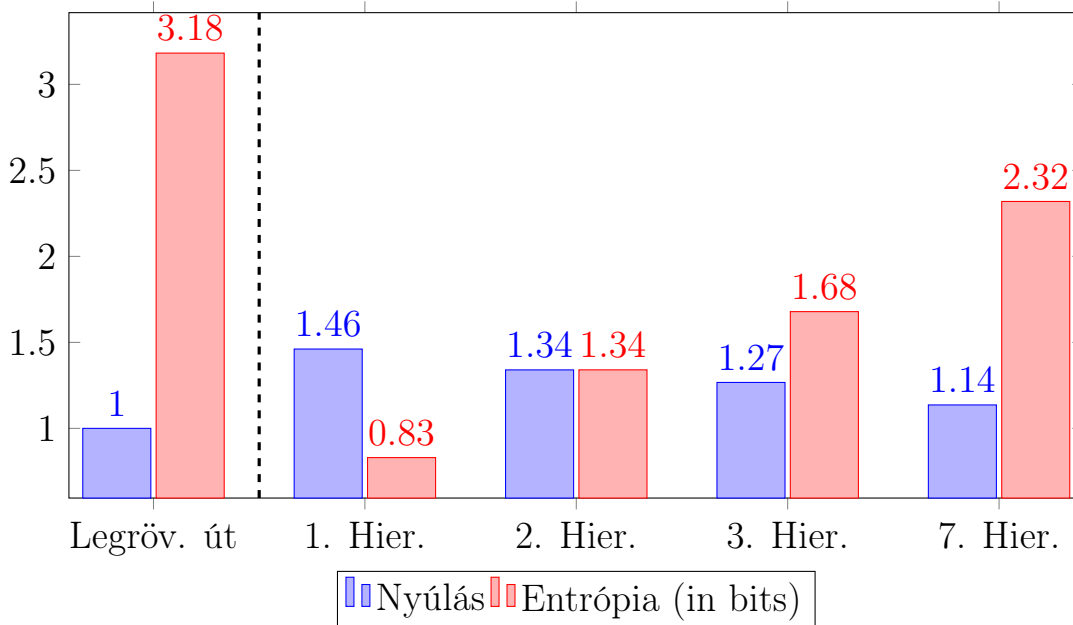


7. ábra. Az emberi és a legrövidebb utak mögötti struktúra szignifikáns különbsége.

ki a hálózattal kapcsolatban. A mögöttes szerkezet nagyon hasonlít egy tiszta hierarchiára. A hierarchia közepe egy rendkívül sokat használt rész, melyen a piros, narancs és zöld színnel kiemelt élet jeleznek. Az ember alanyok tehát, a legrövidebb útválasztó algoritmustól merőben eltérően, nem a teljes gráfon választanak útvonalat, hanem annak egy jóval egy egyszerűbb részgráfja segítségével oldják meg az útválasztás problémáját. Ez a stratégia nyilván szuboptimális, ami egy lehetséges magyarázat az 1.1. Tézisben is megfigyelt útvonalnyúlásra, viszont ennek a stratégiának is vannak kiemelten fontos előnyei.

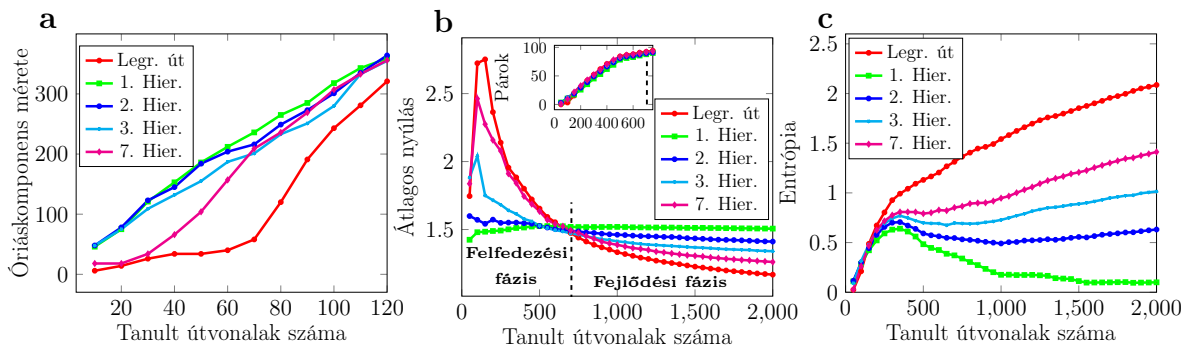
3.3. Tézis ([26, 37, 38]). *Entrópia számítás segítségével megmutattam, hogy a részgráfon alapuló navigáció akár nagyságrenddel is csökkentheti a navigációhoz szükséges memóriaigényt, illetve felgyorsíthatja a hálózati navigáció megtanulásának folyamatát a legrövidebb utakhoz képest.*

A 8. ábrán látható a nyúlás és az entrópia viszonya legrövidebb útválasztás és különféle hierarchikus útválasztási stratégiák esetén. Nyúlás alatt, most a valós útvonal és a legrövidebb út hosszának hányadosát értjük. A legrövidebb útválasztás nyúlása nyilvánvalóan 1, viszont az optimalitás ára a magas entrópia, vagyis a magas memóriaigény. Az 1., 2., 3. és 7. hierarchia mutatja a nyúlás és a memóriaméret közötti kompromisszumos viszonyt szintetikus felépített hierarchiák segítségével. Az



8. ábra. A nyúlás és az entrópia kapcsolata különféle útválasztási mód-szerek esetén.

1. Hierarchia egy egyszerű szélességi bejárásból származó bejárési fa, melyhez tartozó entrópia nagyságrenddel alacsonyabb a legrövidebb úthoz tartozónál. A 2., 3. és 7. hierarchia ebből van származtatva úgy, hogy az 1. hierarchiához hozzáveszünk még 1., 2. illetve 6 a hierarchiában felfelé mutató élet minden csomópontnál. Minél több élet veszünk hozzá a hierarchiához egyre csökken a nyúlás, cserébe folyamatosan nő a memóriaigény.



9. ábra. A szóháló tanulási görbéje különféle útválasztási stratégiák alkalmazása esetén.

Az 9. ábra a hálózati navigáció különböző útvonalakon alapuló el-sajátításának tanulási görbéit tartalmazza. A tanulás elve egyszerűen az, hogy hálózathoz véletlenszerűen kiválasztott mintaútvonalakat mutatunk és azt vizsgáljuk, hogy pusztán a mintaútvonalak ismeretében

milyen navigációs tudást lehet szerezni. Nagyon hasonlít ez ahhoz, ha valaki ismeretlen pl. a fővárosban és egy ismerőse segíti őt az első néhány útján. A kérdés: miután egyedül marad milyen sikeresen tájékozódik a városban. A 9.a ábra a tanult útvonalak számának függvényében mutatja legnagyobb összefüggő komponens méretét, vagyis azon pontokból álló gráf méretét, melyek között önállóan képesek vagyunk közlekedni. Látható, hogy ha a tanító útvonalakat a legrövidebb útvonalak szerint választjuk, akkor a legnagyobb összefüggő komponens mérete nagyon lassan növekszik (lassan áll össze a kép) összehasonlítva a hierarchikus útvonalakkal. A 9.b ábra kis grafikonjáról leolvasható, hogy kb. 700 tanító útvonal ismeretében a pontpárok 90%-a között képesek vagyunk navigálni a tanulómintáktól függetlenül. Ezt a 0-700 tanulási mintáig terjedő szakaszt felfedezési fázisnak nevezzük. A felfedezési fázisban a legrövidebb utakkal való tanítással nem csak az gond, hogy lassan áll össze a kép, hanem a 9.b tanúsága szerint az is, hogy sikeres navigáció esetén is nagy nyúlással kell számolnunk a hierarchikus módszerekhez képest. A felfedezési fázis utáni ún. fejlődési fázisban már lényegében nem tanulunk új pontokat a hálózatból, hanem az eddigi ismereteinket fejlesztjük: jobb útvonalakat használunk. Ebben a fázisban a legrövidebb utakat használó módszer nyúlás tekintetében már megelőzi a hierarchikus módszereket. Tehát a felfedezési fázisban a hierarchikus módszerek hatékonysága jobb a fejlődési fázisban pedig a legrövidebb útválasztásé. A 9.c ábrán a tanulási módszerekhez tartozó entrópia látható. A felfedezési fázisban gyakorlatilag nincs különbség memóriaigény tekintetében, utána viszont jól láthatóan szétválik a különféle módszerekhez szükséges memória.

5. Eredményeim hasznosíthatósága és hatása

A hálózatok funkció→struktúra alapú leírása és megértése gyakorlatilag bármilyen komplex hálózati rendszer esetén hasznosítható az útvonalak és a belőlük következő szerkezeti sajátosságok megértésére, a navigáció javítására (rontására), illetve összeköttetések előrejelzésére. Az Internet esetében fontos problémakör az útválasztók táblázatainak robbanásszerű növekedése [40]. A navigáció jelenségének pontosabb megértése inspirálhat olyan hálózati megoldásokat és protokollokat, melyek az útválasztást sokkal egyszerűbbé és hibákkal szemben ellenállóbbá tehetik. A 2. és 3. tézisekben szereplő hálózati útvonalakkal kapcsolatos méréseim pontosabbá tehetik a valós hálózatokban végbemenő kommunikáció le-

írását, mint a jelenleg elfogadott legrövidebb út hipotézis. Az útvonalak tulajdonságainak pontosabb ismerete elősegítheti társadalmi, kognitív pszichológiai, neurológiai, biológiai valamint ökológiai folyamatok megértését és előrejelezni azok jövőbeli alakulását, de akár gyógyszerek hatásainak és mellékhatásainak analízisében is felhasználható. A 2. és 3. tézisek állításait publikálásuk óta leginkább az agykutatásban hasznosították [41, 42, 43, 44, 45, 46], ahol a valódi mérések nagyon nehézkesek, ezért az analitikusan kezelhető és predikcióra is használható technikák fontossága kiemelt.

Hivatkozások

- [1] Mark EJ Newman. “The structure and function of complex networks”. In: *SIAM review* 45.2 (2003), pp. 167–256 (cit. on p. 1).
- [2] S.N. Dorogovtsev and J.F.F. Mendes. “Evolution of networks”. In: *Advances in physics* 51.4 (2002), pp. 1079–1187 (cit. on p. 1).
- [3] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016 (cit. on p. 1).
- [4] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Vol. 1. Cambridge University Press Cambridge, 2008 (cit. on p. 3).
- [5] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. “Identification of influential spreaders in complex networks”. In: *Nat. Phys.* 6.11 (2010), pp. 888–893 (cit. on p. 3).
- [6] Duncan J Watts, Peter Sheridan Dodds, and Mark EJ Newman. “Identity and search in social networks”. In: *Science* 296.5571 (2002), pp. 1302–1305 (cit. on p. 3).
- [7] Romualdo Pastor-Satorras and Alessandro Vespignani. “Epidemic spreading in scale-free networks”. In: *Phys. Rev. Lett.* 86.14 (2001), p. 3200 (cit. on p. 3).
- [8] Chris J Rhodes and Roy M Anderson. “Power laws governing epidemics in isolated populations”. In: *Nature* 381.6583 (1996), pp. 600–602 (cit. on p. 3).
- [9] Neil Ferguson. “Capturing human behaviour”. In: *Nature* 446.7137 (2007), pp. 733–733 (cit. on p. 3).

- [10] Christian Doerr, Norbert Blenn, and Piet Van Mieghem. “Lognormal infection times of online information spread”. In: *PloS ONE* 8.5 (2013), e64349 (cit. on p. 3).
- [11] Sandro Meloni, Alex Arenas, and Yamir Moreno. “Traffic-driven epidemic spreading in finite-size scale-free networks”. In: *Proc. Natl. Acad. Sci. USA* 106.40 (2009), pp. 16897–16902 (cit. on p. 3).
- [12] Marc Barthélemy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani. “Velocity and hierarchical spread of epidemic outbreaks in scale-free networks”. In: *Phys. Rev. Lett.* 92.17 (2004), p. 178701 (cit. on p. 3).
- [13] Giovanna Miritello, Esteban Moro, and Rubén Lara. “Dynamical strength of social ties in information spreading”. In: *Phys. Rev. E* 83.4 (2011), p. 045102 (cit. on p. 3).
- [14] Lazaros K Gallos, Chaoming Song, Shlomo Havlin, and Hernán A Makse. “Scaling theory of transport in complex biological networks”. In: *Proc. Natl. Acad. Sci. USA* 104.19 (2007), pp. 7746–7751 (cit. on p. 3).
- [15] Yamir Moreno, Maziar Nekovee, and Amalio F Pacheco. “Dynamics of rumor spreading in complex networks”. In: *Phys. Rev. E* 69.6 (2004), p. 066130 (cit. on p. 3).
- [16] Albert-László Barabási and Zoltán N Oltvai. “Network biology: understanding the cell’s functional organization.” In: *Nat. Rev. Genet.* 5.2 (Feb. 2004), pp. 101–13. DOI: 10.1038/nrg1272 (cit. on p. 3).
- [17] Takuji Yamada and Peer Bork. “Evolution of biomolecular networks: lessons from metabolic and protein interactions.” In: *Nat. Rev. Mol. Cell. Bio.* 10.11 (Nov. 2009), pp. 791–803. DOI: 10.1038/nrm2787 (cit. on p. 3).
- [18] E Bullmore and O Sporns. “Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems”. In: *Nat. Rev. Neurosci.* 10 (2009), pp. 168–198. DOI: 10.1038/nrn2575 (cit. on p. 3).
- [19] Dante Chialvo. “Emergent complex neural dynamics”. In: *Nat. Phys.* 6.10 (Oct. 2010), pp. 744–750. DOI: 10.1038/nphys1803 (cit. on p. 3).
- [20] Mark Newman. *Networks: an introduction*. Oxford university press, 2010 (cit. on p. 3).

- [21] Stanley Milgram. “The small world problem”. In: *Psychology today* 2.1 (1967), pp. 60–67 (cit. on p. 3).
- [22] Peter Sheridan Dodds, Roby Muhamad, and Duncan J. Watts. “An Experimental Study of Search in Global Social Networks”. In: *Science* 301.5634 (2003), pp. 827–829. ISSN: 0036-8075. DOI: 10.1126/science.1081058. eprint: <http://science.sciencemag.org/content/301/5634/827.full.pdf>. URL: <http://science.sciencemag.org/content/301/5634/827> (cit. on p. 3).
- [23] Robert West and Jure Leskovec. “Human wayfinding in information networks”. In: *Proceedings of the 21st international conference on World Wide Web*. ACM. 2012, pp. 619–628 (cit. on p. 3).
- [24] Sudarshan Iyengar, Nina Zweig, Abhiram Natarajan, and Veni Madhavan. “A network analysis approach to understand human-wayfinding problem”. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*. Vol. 33. 2011 (cit. on p. 3).
- [32] Fragkiskos Papadopoulos, Maksim Kitsak, M. Ángeles Serrano, Marián Boguñá, and Dmitri Krioukov. “Popularity versus similarity in growing networks”. In: *Nature* 489 (Sept. 2012), pp. 537–540. DOI: 10.1038/nature11459 (cit. on p. 6).
- [33] Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguñá. “Hyperbolic Geometry of Complex Networks”. In: *Phys. Rev. E* 82 (2010), p. 36106. DOI: 10.1103/PhysRevE.82.036106 (cit. on p. 6).
- [34] Richard M Karp. “Reducibility among combinatorial problems”. In: *Complexity of computer computations*. Springer, 1972, pp. 85–103 (cit. on p. 6).
- [35] Geoff Huston. “Analyzing the Internet’s BGP routing table”. In: *The Internet Protocol Journal* 4.1 (2001), pp. 2–15 (cit. on p. 9).
- [40] *{BGP Table Data}*, *{\tt http://bgp.potaroo.net/}*. URL: <http://bgp.potaroo.net/> (cit. on p. 15).
- [41] Danielle S Bassett and Edward T Bullmore. “Small-world brain networks revisited”. In: *The Neuroscientist* 23.5 (2017), pp. 499–516 (cit. on p. 16).
- [42] Andrea Avena-Koenigsberger, Bratislav Misic, and Olaf Sporns. “Communication dynamics in complex brain networks”. In: *Nature Reviews Neuroscience* 19.1 (2018), p. 17 (cit. on p. 16).

- [43] Caio Seguin, Martijn P Van Den Heuvel, and Andrew Zalesky. “Navigation of brain networks”. In: *Proceedings of the National Academy of Sciences* 115.24 (2018), pp. 6297–6302 (cit. on p. 16).
- [44] Alberto Cacciola, Alessandro Muscoloni, Vaibhav Narula, Alessandro Calamuneri, Salvatore Nigro, Emeran A Mayer, Jennifer S Labus, Giuseppe Anastasi, Aldo Quattrone, Angelo Quartarone, et al. “Coalescent embedding in the hyperbolic space unsupervisedly discloses the hidden geometry of the brain”. In: *arXiv preprint arXiv:1705.04192* (2017) (cit. on p. 16).
- [45] Ioannis Pappas, Michael M Craig, David K Menon, and Emmanuel A Stamatakis. “Structural optimality and neurogenetic expression mediate functional dynamics in the human brain”. In: *Human Brain Mapping* (2020) (cit. on p. 16).
- [46] Andrea Avena-Koenigsberger, Xiaoran Yan, Artemy Kolchinsky, Martijn van den Heuvel, Patric Hagmann, and Olaf Sporns. “A spectrum of routing strategies for brain networks”. In: *PLoS computational biology* 15.3 (2019), e1006833 (cit. on p. 16).

Tézisekhez kapcsolódó publikációk

- [25] Attila Csoma, Attila Kőrösi, Gábor Rétvári, Zalán Heszberger, József Biró, Mariann Sliz, Andrea Avena-Koenigsberger, Alessandra Griffa, Patric Hagmann, and András Gulyás. “Routes obey hierarchy in complex networks”. In: *Scientific reports* 7.1 (2017), pp. 1–7 (cit. on pp. 3, 12).
- [26] András Gulyás, József Biró, Gábor Rétvári, Márton Novák, Attila Kőrösi, Mariann Sliz, and Zalan Heszberger. “The role of detours in individual human navigation patterns of complex networks”. In: *Scientific Reports* 10.1 (2020), pp. 1–10 (cit. on pp. 3, 12, 13).
- [27] András Gulyás, József J Biró, Attila Kőrösi, Gábor Rétvári, and Dmitri Krioukov. “Navigable networks as Nash equilibria of navigation games”. In: *Nature communications* 6 (2015) (cit. on pp. 5, 7, 8).
- [28] András Gulyás, Attila Kőrösi, Dávid Szabó, and Gergely Biczók. “On Greedy Network Formation”. In: *Proceedings of the SIGMETRICS W-PIN WS* (2012) (cit. on p. 5).

- [29] András Gulyás, Attila Kőrösi, Gábor Rétvári, József Biró, and Dávid Szabó. “Brief announcement: network formation games can give rise to realistic networks”. In: *Proceedings of the 2012 ACM symposium on Principles of distributed computing*. 2012, pp. 329–330 (cit. on p. 5).
- [30] András Gulyás, Attila Kőrösi, Dávid Szabó, and Gergely Biczók. “On greedy network formation”. In: *ACM SIGMETRICS Performance Evaluation Review* 40.2 (2012), pp. 49–52 (cit. on p. 5).
- [31] David Szabo, Attila Kőrösi, József Biró, and András Gulyás. “Deductive way of reasoning about the internet AS level topology”. In: *Chinese Physics B* 24.11 (2015), p. 118901 (cit. on pp. 5, 11).
- [36] Dávid Szabó and András Gulyás. “Notes on the Topological Consequences of BGP Policy Routing on the Internet AS Topology”. In: *Advances in Communication Networking*. Springer Berlin Heidelberg, 2013, pp. 274–281 (cit. on p. 11).
- [37] Attila Kőrösi, Attila Csoma, Gábor Rétvári, Zalán Heszberger, József Biró, János Tapolcai, István Pelle, Dávid Klajbár, Márton Novák, Valentina Halasi, et al. “A dataset on human navigation strategies in foreign networked systems”. In: *Scientific data* 5 (2018), p. 180037 (cit. on pp. 12, 13).
- [38] Attila Kőrösi, Attila Csoma, Gábor Rétvári, Zalán Heszberger, József Biró, János Tapolcai, István Pelle, Dávid Klajbár, Márton Novák, Valentina Halasi, and András Gulyás. “fit-fat-cat dataset”. In: *Open Science Framework* (2018). eprint: <http://dx.doi.org/10.17605/OSF.IO/JTYVD> (cit. on pp. 12, 13).
- [39] fit-fat-cat. *Smartphone application*. <https://play.google.com/store/apps/details?id=hu.bme.tmit.lendulet.wordnavigationgame>. [Online; accessed 03-20-2019]. 2016 (cit. on p. 12).