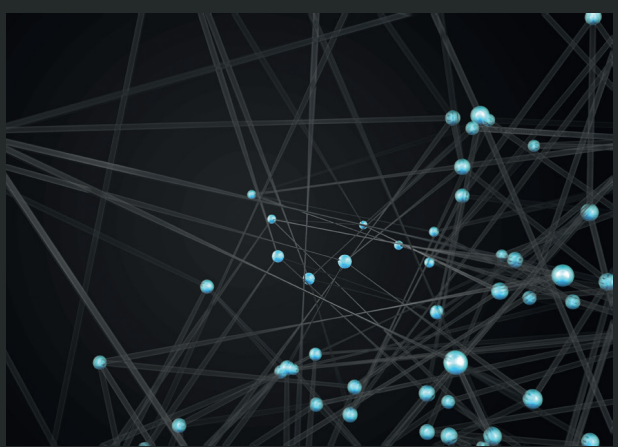
Hálózatok a közszolgálatban



Szerkesztette:
AUER ÁDÁM és JOÓ TAMÁS
Lektorálta:
BARABÁSI ALBERT-LÁSZLÓ

dialóg Campus

Hálózatok a közszolgálatban

HÁLÓZATOK A KÖZSZOLGÁLATBAN

Szerkesztette:
Auer Ádám és Joó Tamás

Dialóg Campus ❖ Budapest, 2019

A kiadvány a KÖFOP-2.1.2-VEKOP-15-2016-00001
„A jó kormányzást megalapozó közszolgálat-fejlesztés”
című projekt keretében jelent meg.

Szerzők:

Auer Ádám
Bányász Péter
Barabási Albert-László
Bederna Zsolt
Dobos László
Gál András Levente
Joó Tamás
Krasznay Csaba
Lőrincz Orsolya
Orbók Ákos
Palicz Tamás
Palla Gergely
Pollner Péter
Ruppert Péter
Sasvári Péter
Szádeczky Tamás
Szalánczi-Orbán Virág
Szócska Miklós
Tibély Gergely
Urbanovics Anna
Váczi Dániel
Varga Melinda

Lektorálta:
Barabási Albert-László

Tartalom

Bevezető gondolatok 7

Szócska Miklós – Joó Tamás – Gál András Levente – Lőrincz Orsolya – Auer Ádám – Palicz Tamás
A hálózatkutatás alkalmazhatósága a közszolgáltatások fejlesztésében 9

Varga Melinda – Ruppert Péter – Barabási Albert-László
Komplex hálózatok szerkezetének elemzése és modellezése 27

Bányász Péter – Dobos László – Palla Gergely – Pollner Péter
Lélektani műveletek a közösségi médiában 111

Krasznay Csaba – Dobos László – Palla Gergely – Pollner Péter
Információbiztonsági incidensek a közigazgatásban 135

Orbók Ákos – Dobos László – Palla Gergely – Pollner Péter
Az egyetemi polgárok wififelhasználói szokásai a Ludovika Campuson 155

Sasvári Péter – Urbanovics Anna – Tibély Gergely – Palla Gergely
Társadalomtudományi doktori iskolák társpublikációs hálózatának elemzése 175

Váczi Dániel – Bederna Zsolt – Szalánczi-Orbán Virág – Szádeczky Tamás
Az incidenskezelés szervezeti háttere 205

Bederna Zsolt – Váczi Dániel – Pollner Péter – Szádeczky Tamás
Támadás hálózatba szervezve 223

Vákát oldal

Bevezető gondolatok

A Nemzeti Közszolgálati Egyetem 2015–2020 közötti Intézményfejlesztési Tervének jelmondata, hogy az NKE „Az együttműködés egyeteme”. Őszintén bízunk benne, hogy ez a könyv jelentős lépés e célkitűzés megvalósításához.

2017 tavaszán az Egyetem a Digitális Jólét Program részeként közreműködött Magyarország Hálózatkutatási Stratégiájának megalkotásában. A programelem ötletgazdájaként a Digitális Jólét Program és az NKE vezetése úgy gondolta, hogy tartós eredményt abban az esetben lehet elérni, amennyiben a projektnek primer kutatási eredményei és ebből kifejlesztett oktatási kapcsolódása, tananyaga is lesz. Az akkori Államkutatási és Fejlesztési Intézet szervezésében alakult ki egy együttműködési modell, amelynek egyik eredménye ez a kötet. De hogyan jutottunk el idáig?

A hálózatkutatás korunk egyik legkurrensebb kutatási területe, amelynek alkalmazása számos területen még csak ötlet szintjén található meg. A hálózatkutatás egyik alapvető eleme az adat, amelyen vagy amellyel kutatásokat lehet folytatni. De önmagában ez még nem elégséges az eredmények eléréséhez, hanem hálózattudományi ismeretek is szükségesek hozzá. Ugyanakkor még talán ezen kettős elemi feltétel megléte esetén sem várhatunk eredményeket. Szükséges az a szakmai háttér, amely megfogalmazza a hipotéziseket, segít eligazodni az adatok által kirajzolt mintákon, eredményeken, és végső soron hasznosítani tudja ennek az eredményeit.

A projekt szoros együttműködési modellt feltételezett: egyrészt a kutatás tárgyához kapcsolódó szakmai oldal, valamint a hálózattudományi kutatók együttműködésének köszönhetően juthattunk el oda, hogy a kutatások eredményesek legyenek. A projekt egyik fő mentora Barabási Albert-László, aki az ötlet felmerülése óta közreműködött a különböző fázisokban. A kutatás nyitórendezvényén, 2018. március 21-én hallhatta a közönség, hogy a hálózattudomány milyen kihívásokkal küzd manapság. Ezt követően indult el egy hosszabb munka, amelynek célja olyan eredmények elérése volt, mint amelyeket dr. Szócska Miklós és szakmai csapata korábban, a szakigazgatás egy területén már elért. Úgy gondoltuk, hogy ezen az úton végighaladva be tudjuk mutatni azokat a lehetséges irányokat, amelyek későbbi kutatások alapjai lehetnek.

A kötetben szereplő tanulmányok mögött álló kutatások eredményeit 2018. decemberben tudományos konferencián mutattuk be. Ezen a konferencián a hallgatóság számára nyilvánvalóvá vált, hogy a hálózatkutatás alkalmazott tudományos eredményeit milyen széles körben lehet(ne) beépíteni a napi gyakorlatba.

A kötet célja egyrészt, hogy a közigazgatás szerteágazó területeiről (tudatosan ilyen széles a spektrum, és nem fókuszál egy-egy kérdésre) mutasson be esettanulmányokat. Így konkrét problémák megoldását lehet modellezni, és szemléltethető a hálózatkutatási módszer alkalmazása. Ez alkalmas arra, hogy közigazgatási szakemberek megismerkedjenek ezzel a területtel. Másrészt a tanulmányok célja az is, hogy az alapvető hálózatkutatási

fogalmakkal megismertessék az olvasót. Aki a kötetet végigolvassa és alaposan áttanulmányozza, a hálózatkutatási módszer elemeit és az eredmények értelmezését is elsajátíthatja. Nem titkolt célja a kötetnek, hogy egy alkalmazott hálózatkutatási tantárgy oktatási segédanyaga is legyen, így ha esetlegesen indulna ilyen kurzus, akkor szemléltethető, példákkal illusztrált és lektorált tananyagot tudunk a hallgatók rendelkezésére bocsátani.

Az együttműködéssel kezdődött ez a bevezető, és ez egyáltalán nem elhanyagolható szempontja ennek a projektnek. Egyrészt a Digitális Jólét Programmal való együttműködés tette lehetővé ezeket a kutatásokat, másrészt kiemelkedő közös kutatómunka folyt a Semmelweis Egyetem, az Eötvös Loránd Tudományegyetem és a Nemzeti Közszolgálati Egyetem kutatói között. „Az együttműködés egyetemeként" az NKE-n folytatott kutatómunkában sikerült egy olyan partnerséget megalapozni, amely remélhetőleg nem egyszeri, hanem tartós eredményeket érhet el.

Budapest, 2019. június 10.

Auer Ádám

Szócska Miklós – Joó Tamás – Gál András Levente –
Lőrincz Orsolya – Auer Ádám – Palicz Tamás

A hálózatkutatás alkalmazhatósága a közszolgáltatások fejlesztésében

Bevezetés

A modern világ egymással növekvő kölcsönhatásban lévő kiber-, társadalmi, technikai, természeti, környezeti hálózatok gyűjteményére támaszkodik. Minden egyes hálózat önmagában összetett, meghatározott jellemzőkkel és a szabadság számos, igen különböző fokával rendelkezik. A hálózattudomány egy erőteljes koncepcionális és gyakorlati keretet biztosít ahhoz, hogy értékeljük és modellezzük ezeket az összetett természeti, technológiai és társadalmi rendszereket. A hálózati minta egyedülálló lehetőséget biztosít arra is, hogy időn és téren át integráljuk az elemek különböző típusait (társadalmi, technológiai, természeti stb.).

A hálózattudomány fejlődése számos, együttesen megjelenő tényezőnek köszönhető: a sokrétű és dinamikus hálózatokhoz szükséges fejlettebb matematikai képességek; nagy tömegű humán adathoz történő hozzáférés, beleértve az egészséget, a mobilitást és a kommunikációt; nagy teljesítményű számítógépes technológia a nagy adatrendszerek rögzítésére és kezelésére; valamint – általánosabb megközelítésben – a hálózati gondolkodás elfogadása az egyetemek, a kormányok és az ipar részéről.

Ennek eredményeként az elmúlt két évtizedben a hálózatokon alapuló eszközök jelentőssé váltak az összetett rendszerek mechanizmusainak és elveinek felfedezésében, és olyan jelenségek feltárásához járultak hozzá, amelyek korábban más megközelítésekkel nem voltak érzékelhetők.[1]

A hálózatkutatás eddigi közigazgatási alkalmazási területei

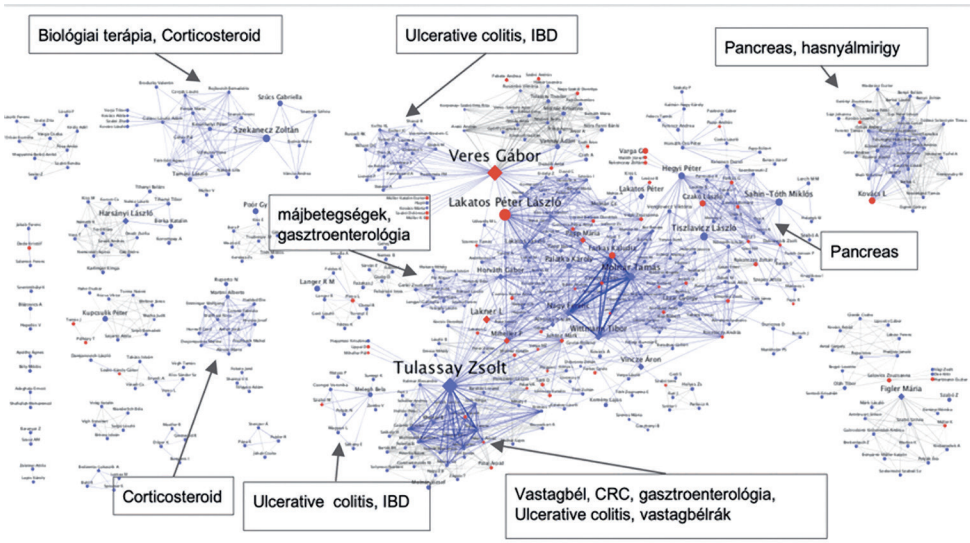
A hálózatkutatás elsődleges választás lehet azokban az esetekben, amikor egy rendszer különböző szereplői, tényezői vagy éppen eseményei közötti összefüggéseket szeretnénk leírni jól mérhető adatok alapján, valamint abban az esetben is hasznos, amikor a kevésbé transzparens összefüggések feltárása a cél. A hálózatelemzés segítségével tehát lehetőség nyílik direkt és indirekt, látható és „láthatatlan" összefüggések feltárására nagy mennyiségű, strukturálatlan adat esetén is. Fontos megjegyezni, hogy a hálózatelemzési módszerek nem csak egyszerű

[1] Barabási 2017.

kapcsolatelemzésre szolgálnak: adott területen sokkal komplexebb és kifinomultabb, informális tudást biztosíthatnak (például a piaci és szervezeti struktúrákra vonatkozóan).

A hálózatelemzés már széles körben ismert alkalmazási területeire néhány példa:

- Publikációs adatbázisok elemzésével „tudáshálózatok" elemzése különböző tudományterületeken és ez alapján innovációs kutatások összehangolt támogatása (lásd az 1. ábrát).
- Szervezeti információkat (például cégregiszter) felhasználva K+F-befektetések hatásának és diffúziójának elemzése, valamint ez alapján a piaci szereplők célorientált ösztönzési programjainak kidolgozása.
- Menedzsmentfókuszú piaci beavatkozások hatásának mérése és ez alapján transzparens támogatási rendszer működtetése.
- Pályázati adatok és nyertes szervezetekre vonatkozó információk elemzése és ez alapján a közbeszerzési folyamatok vizsgálata, trendek és mintázatok kimutatása, partnerkiválasztási stratégia.
- Szakmai vagy piaci véleményvezérek, befolyásolók azonosítása.



1. ábra

Publikációs adatbázisok elemzése hálózatos módszerekkel

Megjegyzés: a hálózati ábrán a magyar gasztroenterológusok (belgyógyász és sebész szakorvosok) publikációs hálózata és szakterületi koncentrációi láthatók.

Forrás: Fiatal Gasztroenterológusok Munkacsoportjának 10. jubileumi kongresszusa, Szócska Miklós előadása

Egy adott kutatási terület tudáshálózata jól modellezhető például releváns publikációs adatbázisok feldolgozásával, elemzésével. A tudáshálózat modellezésével választ kapunk arra, hogy kik tekinthetők egy-egy szakterület mérvadó szakértőinek, illetve kik azok a szereplők, akiknek a szakmai „értéke" várhatóan növekedni fog a jövőben. Tehát azonosíthatók a vé-

leményvezérek és a feltörekvő szakértők. Egy ilyen szakmai-tudományos tudáshálózat feltérképezése számtalan gyakorlati lehetőséget rejt. Könnyebb, célzottabb és gyorsabb lehet például új, innovatív terápiák bevezetése és elterjesztése. Növelhető az adott szakterületen az ellátásokhoz való hozzáférés, az ellátás minősége, a betegbiztonság.

A tudáshálózatok analizálása számos területen felhasználható. Alább a teljesség igénye nélkül néhány egészségügyből vett példát sorolunk fel, amelyek analógiájára más közszolgálati ágazatok problémái is elemezhetők, kezelhetők:

- Kapacitások tervezése: az adott intézményhez, területhez kötődő szakemberek száma, kompetenciái, az általuk elérhető ellátási teljesítmény.
- Menedzsmentkontroll, a finanszírozási beavatkozások célzott megvalósítása; transzparencia, szakmai érdekek és formális-informális üzleti modellek feltárása (a magán- és közszolgálati ellátás láthatatlan összefüggéseinek feltárása, bizonyítása).
- Klinikai kutatások: az adott helyen elérhető szakmai kompetenciák, a jelenleg futó programokra lekötött és szabad vizsgálói kapacitások.
- A tudáshálózatok természetesen nemcsak egy-egy szakterületen vizsgálhatók (például kulcsszavak vagy személyek szerint), hanem intézményhez, országhoz, K+F és innovációs pályázati forrásokhoz, szerzőcsoportokhoz kötődően is felépíthetők.

Hálózatos ismeretek szerepe a közintézményekben

A közintézmények vezetőinek és munkatársaiknak egyre összetettebb társadalmi és politikai összefüggések mentén kell hatékony és jól tervezhető döntéseket hozniuk vagy ilyen döntések előkészítésében segédkezniük. Ugyanakkor az egyre jobban hozzáférhető számítógépes adatbázisok, nyilvántartások és ügykezelői rendszerek nagy mennyiségben állítanak elő háttérinformációkat. Ezen információk megértéséhez, átlátható kezeléséhez többnyire szakértői programok vagy egyéni tanácsadók adnak elemzéseket, értékeléseket. Azonban a szoftverek kezelése, a tanácsadók értékeléseinek gyakorlatba való átültetése akkor lesz igazán hatékony, ha a döntéshozók, de akár a hétköznapi ügyeket intéző köztisztviselők is rendelkeznek a megfelelő alapfogalmak és összefüggések ismeretével az adat- és hálózattudományok területén.

Mivel az elemzésekhez szükséges számítási kapacitások könnyen elérhetők, ezért a hálózatelemzési képességek kiépítéséhez manapság csak három dologra van szükség:

1. adatbázisokra és az azokhoz való jogtiszta hozzáférésre,
2. a hálózatelemzési, adattudományi módszertant ismerő szakemberekre
3. és végül, de egyáltalán nem utolsósorban olyan ágazatismerettel rendelkező szereplőkre, akik a megfelelő elemzési kérdéseket meg tudják fogalmazni a szakemberek felé, valamint értelmezni tudják az érintettek számára az elemzéseket.

A hálózatos szemlélet kialakítása nem más, mint azoknak a képességeknek a megteremtése, amelyekkel összetett folyamatokat és többlépéses feladatokat is hatékonyan át lehet látni, meg lehet tervezni. Különösen azon eseteknél hasznos ez az újszerű szemlélet, ahol több lehetőség irányába is el lehet indulni. Azokban a kérdésekben, ahol több reláció is meghatározza a szereplők feladatait, előnyös lehet a hálózati csúcsok és élek megfelelően elrendezett értelmezése.

A modern állam emberi hálózatok összehangolt együttműködésével képes csak hatékonyan működni. Az együtt dolgozó alkalmazottak és vezető tisztségviselők akkor tudják

munkájukat megfelelően koordinálni, ha az egymással való kommunikáció során azonos alapfogalmakat használnak.

Az objektív döntéshozatalt és döntés-előkészítő munkát számszerűen mérhető jellemzők segítségével célszerű megtervezni. A döntést befolyásoló gazdasági, társadalmi vagy politikai szempontokat tanácsadói szinten mélyebb matematikai, statisztikai, valószínűségszámítási kalkulációkkal számszerűsítik. Ezen elemzések eredményeként számos hálózati ismérvet lehet kimutatni. Azok a tisztségviselők és döntéshozók, akik ismerik e jellemzők alapvető tulajdonságait, sokkal nagyobb mértékben tudják az állampolgárok javára fordítani a hatástanulmányok, illetve döntés-előkészítő elemzések eredményeit.

Az állami feladatok sokszor igényelnek olyan tervezési és modellezési képességeket, amelyek a jelenlegi és múltbeli adatok alapján a jövőben várható változásokról átfogó kép kialakítását, megfogalmazását teszik lehetővé. Az is fontos szempont, hogy a szakértők által elkészített elemzéseket és előrejelzéseket proaktív módon tudják kezelni, tehát egy-egy modell eredménye ne csak egy feketedobozból előhúzott varázslat legyen a számukra. Az egy-egy szakterületre vagy specifikus kérdésre koncentráló szakértő ugyanis más területről vagy tágabb összefüggésrendszerből származó hatásokat esetleg nem vesz figyelembe, vagy nincs birtokában ezeknek az információknak. A modellezési eredményeket értő módon átvevő vezető képes lesz érzékelni az ilyen hiányosságokat, és megfelelő intézkedésekkel vagy a szakértők részletesebb tájékoztatásával kerülheti el az esetleges tévedéseket.

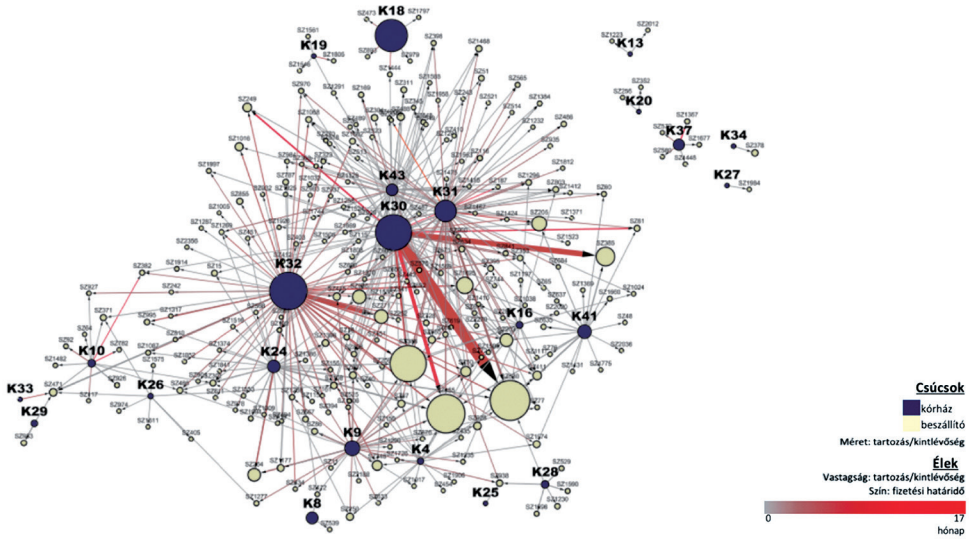
A vezetői és döntéshozói munka együtt jár bizonyos átalakítási vagy átszervezési intézkedésekkel, új egységek vagy új kapcsolatok létrehozásával, esetleg a régiek megszüntetésével. Ezekben az intézkedésekben hasznosulhatnak olyan ismeretek, amelyek az összetett, többszörös kapcsolati viszonyrendszerek robusztusságát írják le. Egyszerű visszajelzési mutatók alapján (például a megszüntetendő egységek száma, kialakítandó kapcsolatok költsége stb.) nehéz belátni, hogy mennyire lesz hatékony a döntés, milyen messzebbre ható következményekkel járhat. Hálózatos szemléletű tervezéssel viszont azonos költséggel vagy azonos átszervezési volumennel akár a teljes viszonyrendszer is átrendezhető, ha a megfelelő kulcspozíciókat sikerül megtalálni.

A hálózatelemzéshez kapcsolódó hálózatos vizualizáció nemcsak a célzott beavatkozások lehetőségét teremti meg, hanem intuitív, direkt módon segít hozzá komplex (például üzleti) modellek közvetlen megértéséhez.

Az államháztartás központi alrendszerébe tartozó költségvetési szervek év végi tartozásállománya az elmúlt években elérte a 60–80 milliárd forintot. Ennek jelentős része, 70–80%-a az egészségügyi intézményeknél halmozódott fel (beleértve a járó- és fekvőbeteg-szakellátás intézményeit, az egészségügyi ágazati háttérintézményeket, valamint a klinikai központtal rendelkező felsőoktatási intézményeket). Az elemzés segítségével célzott adósságcsökkentési és központosított országos közbeszerzési rendszerekkel lehetett éves szinten 10 milliárd forint nagyságrendű megtakarítást elérni.

Az államtudományokhoz kapcsolódóan jelentős fejlesztő és problémamegoldó eredményeket lehetne tudományos alapon elérni például az alábbi területeken:

- a rendészettudományban a modern bűnüldözési technikák terén;
- a hadtudományban a hibrid hadviselésben;
- a közigazgatás-tudományban az adóeljárások kapcsán, az elektronikus közigazgatásban, valamint a szervezeti hatékonyság, döntés-előkészítés területein.



2. ábra

Az államháztartás központi alrendszerébe tartozó fekvőbeteg-szakellátás intézményeinek tartozásállománya

Megjegyzés: A hálózat középpontjában a legtöbb tartozással a K30, K31 és K32 azonosítójú kórházak és a hozzájuk kapcsolódó, legnagyobb kintlévőségű beszállítók állnak. Ezek a kapcsolatok nagyrészt már több hónapos lejáratú tartozásokat jelölnek.

Forrás: Magyar Kórházszövetség 2014. évi Kongresszusa, Szócska Miklós előadása

Külön kérdésként jelentkezik a szabályozási környezet (adatvédelem, rendszerbiztonság) kialakításának szükségessége, amely az állam szabályalkotó feladatát érinti.

Összefoglalva: a hálózatos alapon is szerveződő modern társadalmak megfelelő működtetéséhez szükség van hálózatkutatási kompetenciákra. Ez olyan ismereteket és gyakorlati képességeket jelent, amelyekkel a hálózatos tudást egyszerű eszközként lehet használni ismeretszerzésre, problémamegoldásra és döntéshozatalra. Fontos, hogy ez a tudás társadalmi szinten továbbadható legyen, és a hálózatos gondolkodás lehetséges előnyeit és buktatóit az átlagember számára is érthető módon meg tudják fogalmazni.

Egyéni és csoportos hálózatkutatási kompetenciák

Magyarországon a hálózatkutatás fogalma a kétezres évektől lett egyre ismertebb. A viszonylag szűkebb szakértői körön (szociológusok, matematikus, fizikusok) kívül általában nem volt jellemző ezen ismeretek megléte, az alkalmazása pedig különösen hiányzott a döntéselőkészítési eszköztárból. Az elmúlt években a fogalom egyre ismertebb lett, illetve ha nem is mindennapossá, de egyre gyakoribbá vált az elemzési módszertanok között. A szervezetek vezetői egyre több esetben ismerik fel, hogy a hálózatkutatás módszertana egy másféle megközelítést tesz lehetővé, és egyre bátrabban alkalmazzák.

A magyar közigazgatásban a 2010-es időszakot követően jelent meg a hálózattudomány elvein alapuló megközelítés. Ebben jelentős szerepet játszott, hogy új, innovatív szemlélettel rendelkező vezetők jelentek meg, másrészt a 2010-es kormányváltást követően indult el a közigazgatás reformja, a Magyary Zoltán Közigazgatás-fejlesztési Program, amely a korábbiakhoz képest sokkal korszerűbb szemléletmóddal indította el a közigazgatás megújulását. Ennek keretében lehetővé vált új elemzési módszerek kipróbálása is. Azonban jelenleg is több tényező akadályozza, hogy e módszerek széles körben elterjedjenek:

- A közigazgatásban dolgozók körében továbbra sem kellőképpen elterjedt a hálózattudomány, a hálózatos gondolkodás, illetve az ezzel nyerhető információk köre.
- A közigazgatásban dolgozó vezetők ismeretei és attitűdje nem teremt megfelelő környezetet a döntés-előkészítési és döntési folyamatokban történő alkalmazásra.
- A jelenlegi oktatási-képzési rendszer nem biztosítja a hálózatkutatási ismeretek hasznosítását, alkalmazását.
- A közigazgatás továbbképzési rendszerében számos hiányosság van.
- A magyar akadémiai közeg nem rendelkezik akkora kapacitással, amely biztosítani tudná akár az oktatás kiszélesítését, akár a szakértői tevékenység fokozását.

Ha szélesebb körben vizsgáljuk a közszolgálatások területén dolgozókat, a fentiekhez hasonlókat állapíthatunk meg. Jelenleg csak elszigetelten dolgozó vezetők, műhelyek, iskolák vagy képzések vannak, ahol használják a hálózattudományi ismereteket.

Összefoglalva azt mondhatjuk, hogy a jelenlegi helyzetben az egyéni és a csoportos hálózattudományi ismeretek és kompetenciák jelentősen elmaradnak attól, mint ami a mai modern, hálózatosan szerveződő világ megértéséhez szükséges lenne. Az ehhez szükséges kapacitások, tudásbázisok, módszertani háttérintézmények hiányoznak a jelenlegi magyar közigazgatásból.

Társadalmi és technikai hálózatkutatási infrastruktúrák – nemzeti adatbázisok és adatkataszter

Egy friss adatipari elemzés szerint „az európai adatgazdaság értéke 300 milliárd eurót tesz ki, amely megfelelő jogalkotási és szakpolitikai intézkedések bevezetésével 2020-ra akár 739 milliárd euróra emelkedhet”.[2] Több iparágban az adatok jelentik a technológiai és gazdasági fejlődés, a munkahelyteremtés egyik kiemelkedő lehetőségét. Nem túlzás azt állítani, hogy megkezdődött egy adatalapú paradigmaváltás, az adatvezérelt megoldásokra való átállás Magyarországon, több területen is (például egészségügy). Egyes vezető kutatóintézetek nem egyszerűen a kutatási, fejlesztési és innovációs célú adathasználat lehetőségeit keresik, hanem már egy adatvezérelt paradigmaváltás feltételrendszerének összeállításán dolgoznak.

A technológiai forradalom fénysebességgel zajlik, a versengő új és helyettesítő technológiák még nem letisztultak, azonban most dől el az, hogy ebben a digitális transzformációban Magyarország milyen helyet foglal majd el: hogy hazánk időben lép-e, és ki tudja-e használni a relatív adatgazdagságában és a több területen meglévő, országosan integrált

[2] Az Európai Bizottság sajtóközleménye: *Adatok az EU-ban: a Bizottság fokozza az egészségügyi adatok rendelkezésre állásának növelése és a megosztásuk ösztönzése érdekében tett erőfeszítéseket.* Brüsszel, 2018. április 25.

alapstruktúrájában rejlő lehetőségeit. Mindez azon is múlik, hogy stratégikusan mozgósítani tudja-e a területen meglévő kivételes, de szigetszerű erőforrásait.

Magyarország egyéb technológiafejlesztési prioritásai is kapcsolódnak a digitális forradalomhoz. Például a telekommunikációban aktuális 5G-fejlesztések nemcsak az autonóm autók, hanem a digitális megoldások nagy mennyiségű adatmozgását is lehetővé teszik. Emellett a mesterséges intelligenciához kapcsolódó innovációs prioritások is könnyen csatlakozhatnak más iparági megoldások fejlesztéséhez. Az egészségügyben már most elérhetők azok a mélytanuló mesterségesintelligencia-algoritmusok, amelyek életmentésre vagy hatékonyságjavításra is felhasználhatóvá teszik az egészségügyi adatok megközelítőleg 80%-át kitevő strukturálatlan szöveges és képi információkat. Ezek első, meglévő magyar algoritmusra épülő, jóformán azonnali alkalmazása lehet például az emlőtumorszűrő képalkotó felvételek minőségbiztosítása és a nem kiszűrt tumoros esetek gépi azonosítása. Ez az algoritmus nemzetközi szintű piaci potenciállal rendelkezik.[3]

Az elmúlt 15 év során a hálózattudomány szempontjából az infrastruktúrákat széleskörűen tanulmányozták, ami lehetőséget teremt a különféle rendszerek szerkezetének egységes leírására olyan csúcsok hálózataként (rendszerelemek, például elektromos buszok vagy metróállomások), amelyeket linkek kötnek össze – ezek képviselik a kapcsolatot a csúcsok között (például átviteli vonalak vagy vasutak).

Több jelentős tanulmány foglalkozott például a regionális vagy városi vízellátás hálózati modelljeivel, különös tekintettel arra, hogy azok képesek-e továbbra is kielégíteni a fogyasztói igényeket a változó piaci feltételek, az adókivetések, a növekvő számú lakosság, a természeti katasztrófák, valamint a klímaváltozás körülményei között.

Az energiarendszerekben az előbb megfogalmazottak „intelligens áramhálózatot" jelentenek, amely zavarok esetén képes az öngyógyításra és intelligens válasz adására, és amelynek szilárdan ellen kell állnia a generátorszinkron megszűnése vagy a feszültség összeomlása esetén (mindkettő fontos bűnös az áramszünet előidézésében).

A közlekedési rendszerekben egy ilyen rendszerrel elkerülhető lenne a torlódás, és ez elősegítené az emberek és áruk szabad áramlását természeti katasztrófák vagy terrorista támadás esetén is. Az e területen tett erőfeszítések többségükben a rendszer strukturális és technológiai integritására és összekapcsolhatóságára koncentrálnak. Azonban az infrastruktúra nem csupán műszaki probléma. Például a közlekedés megbénulhat akkor is, ha egy természeti katasztrófa a főbb utakat, hidakat és vasutakat viszonylag sértetlenül hagyja. Egyre inkább elismerik, hogy ha a hálózati infrastruktúrák szilárdságát és készenlétét valóban kezelni akarjuk, akkor mérnünk kell a használatukat mozgató társadalmi tényezőket is.

E kutatások alapvető kérdései többek között:

- Milyen módon irányíthatjuk a viselkedést a társadalmi tartományban a technikai struktúrába történő beavatkozással?
- Hogyan befolyásoljuk a technikai tartomány hatékonyságát a társadalmi tartomány hálózati struktúráiba történő beavatkozással?

Ezen alapvető kérdések megválaszolása során modelleznünk kell a társadalmi, technikai rendszerek ilyen új kereteinek biztonsági, magánéleti és etikai szempontjait. A társadalmi,

[3] Ribli et al. 2018.

technikai rendszerek komplex hálózatai felé történő elmozdulás mozgatóerejét két központi trend alkotja:

1. a technológiai rendszerek növekvő autonómiája egy kevert rendszert hoz létre, amelyben a hálózat csúcsai lehetnek emberek és más autonóm tényezők is;
2. a közösségi gazdaság (share economy, sharing economy) pedig az emberek és szervezetek közötti társközi (peer-to-peer), dinamikus forrásmegosztás erős hálózatát hozza létre.

A hálózattudomány felhasználható arra, hogy elemezzen, tervezzen és irányítson ilyen összetett társadalmi, technikai rendszereket. Alkalmas a társadalmi, a szenzoros (sensing) és a technológiai tartományok együttes fejlődésének modellezésére, valamint a többrétegű, többforrású hálózatok számára megfelelő elméleti és módszertani keretek kifejlesztésére. Továbbá a társadalmi, technikai rendszerek közösségi viselkedésre gyakorolt hatását irányítani lehet a hálózat struktúráján keresztül.

Végezetül, fontos lesz meghatározni a rendszer fejlődésének pályáját. A társadalmi, technikai rendszerek dinamikájának megértése megköveteli, hogy megértsük az adott rendszer tényezőinek közbenső állapotait, valamint e közbenső állapotok becslése során ismerni kell a fejlődési előrejelzéseket és a bizalmi tőkét. A társadalmi, technikai rendszerek tervezése vagy a társadalompolitikai rendszerek szakpolitikai elemzése esetén szükség lesz egy megfelelően méretezhető és megbízható laboratóriumra, amely képes lesz olyan nagy léptékű kísérletek végzésére, amelyek azután átvihetők lesznek a való világ problémáira. Ennek fontos feltétele olyan módszertani eszköztár kidolgozása, amely lehetővé teszi a megbízható, adatok által irányított szimulációt. Ezt a szimulációs környezetet lehet majd felhasználni a különféle beavatkozások tesztelésére egy sor „mi történik akkor, ha” forgatókönyv alkalmazásával. Egy virtuális kísérleti környezetben megjelenik a replikáció és a reprodukálhatóság kihívása is. Ily módon a szimulációs forgatókönyvekre vonatkozó új jóváhagyási és hitelesítési módszerekre van szükség.

A következő 10-20 évben a közösségi gazdaság, valamint a humán és autonóm tényezők kevert rendszerei irányába történő átalakulás megvalósítása során virtuális, kísérleti szimulációs forgatókönyvek alkalmazása várható az infrastruktúrák tervezésének és irányításának fő pilléreiként.

Tudomány, gazdaság és állam szereplőinek hálózatos együttműködése

Az alábbiakban bemutatjuk azokat a főbb területeket, amelyek esetében a hálózattudomány és kapcsolódó területei hozzájárultak egy-egy szakpolitika vagy nemzeti cél megoldásához. Nemzetközi szinten számtalan, hazai feltételek között is alkalmazható hálózatkutatási elmélet és gyakorlati projekt született már bizonyos államigazgatási kompetenciákat is érintő, komplex társadalmi, gazdasági, politikai kérdések megválaszolására. Az elméleteket részletező publikációk bemutatják, hogy a hálózatkutatás mely módszertani eszközeivel lehet a gyakorlatban egy adott nemzeti vagy helyi szintű, gazdasági vagy politikai kérdést megválaszolni. Az alábbiakban néhány releváns elmélet felsorolása következik, amelyek hálózattudományi alapon adnak választ az egyes szakterületeken felmerülő, összetett kérdésekre.

Közigazgatási hasznosíthatóság: szervezeti hálózatkutatás

A hálózattudomány kiválóan alkalmas szervezeteken belüli vagy szervezetek közötti együttműködési formák mennyiségi és minőségi elemzésére a szervezetek egészére, részére vagy akár egy-egy szervezeti egység szintjére lebontva a vizsgálatot. A szervezeti hálózatkutatás a közigazgatás valamennyi szintjén hasznosítható, hiszen ezzel az eszköztárral a szervezeti szinttől függetlenül az egyes dimenziókban egyfajta röntgenkép készíthető, legyen szó minisztériumról, háttérintézményről vagy intézményi csoportról.

Közigazgatási hasznosíthatóság: kiemelt adatgazda-szervezetek adatkincsének feltárása

Magyarországon számos olyan állami szervezet működik, amelyek adatkincse még nincs teljes mélységben kiaknázva. E szervezetek működése során szinte felbecsülhetetlen értékű adatállomány képződik. Ezt az adott ágazat, illetve országos stratégiai kérdések megválaszolására lehet alkalmazni megfelelő gyűjtés és rendszerezés mellett. A területen a legfőbb szervezeti szereplők a következők:

- Nemzeti Adó- és Vámhivatal,
- Magyar Államkincstár,
- Központi Statisztikai Hivatal,
- Nemzeti Egészségbiztosítási Alapkezelő,
- Állami Egészségügyi Ellátó Központ,
- Oktatási Hivatal.

Az uniós források külön vizsgálati területet jelentenek. Az itt végzett hálózati kutatások a következő kérdésekre adhatnak választ:

- a források hasznosításában részt vevő legfőbb szervezetek hálózata, a közöttük kialakult együttműködés mélysége és rendszeressége;
- a források felhasználása közben létrejött tudástranszfer, többek között az impaktfaktor szintjén mérve.

Közigazgatási hasznosíthatóság: szakpolitikai döntéstámogatás

A hálózati tudományok alkalmazása a magyar közigazgatásban, illetve a fentebb felsorolt területeken nem rendszerszerűen, hanem sporadikusan történt. Ez általában az adott ágazatot, területet vagy intézményt irányító vezető egyéni ismereteire és ambíciójára, proaktivitására vezethető vissza. Az elemzések elkészültét követően viszonylag ritkán valósult meg az elemzéseken alapuló, tudatos beavatkozástervezés, illetve azt követően az elért eredmények kiértékelése.

A hazai környezetben a 2010–2014 közötti időszakon belül jellemzően az egészségügyi ágazatban alkalmazták a hálózati kutatásokat, és ehhez szorosan kapcsolódóan a döntés-előkészítésben és a kommunikációban is felhasználták az így kapott eredményeket. Az ágazatvezetés a hálózatkutatás megfelelő adatvizualizációs eszköztárát is igénybe vette.

Ezen időszak eredményeit áttekintve fontos kiemelni azt a tényt is, hogy sok esetben nem strukturált, egymással kapcsolatban lévő, előre meghatározott adatbázisok elemzése történt, hanem leggyakrabban az ágazaton belül szigetszerűen meglévő adatok összekapcsolásával és azok elemzésével lehetett jelentős eredményeket elérni. A vizsgálat rávilágított arra is, hogy relatíve alacsony költséggel előállíthatók azok az elemzések, amelyek ágazatvezetői szempontból kritikusak, és megfelelő szakmai értelmezéssel és tudással jól használhatók az egyes döntések meghozatalakor.

Közigazgatási hasznosíthatóság: menedzsmentkontroll-eszköz

A komplex rendszerek megértésének és kontrolljának egyik fontos lehetősége a hálózatkutatási eszköztár használata, hiszen ezzel lehetővé válik rejtett összefüggések felismerése is. Ilyen komplex rendszernek tekinthető egy adott ágazat is. Ebben a példában a kórházi konszolidáció előkészítésében használt elemzést mutattuk be, amelyet az egészségügyi ágazatirányítás vett igénybe Magyarországon 2010–2014 között (lásd a 2. ábrát).

Biztonságpolitika: terrorista szerveződések azonosítása

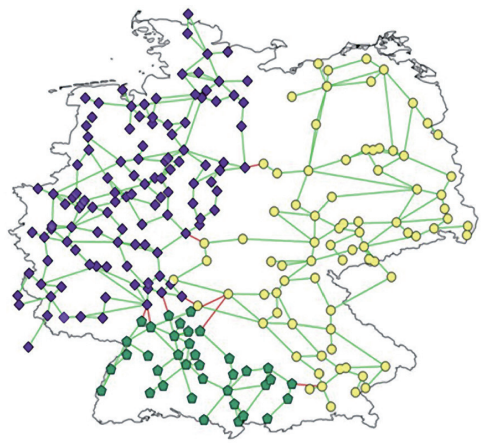
Az online közösségi média adatgyűjtésben eleve gazdag digitális környezete miatt kifejezetten jó feltételeket biztosít hálózati kutatások számára. Az alábbiakban szerepel néhány olyan terület, ahol már Magyarországon is történtek hálózati kutatások egy adott kommunikációs trend elleni pozitív kommunikációs küzdelem kezdeményezésére. Szélsőséges és terrorista csoportok rendszeresen használják az internetet pszichológiai hadviselés, propaganda és megtévesztés céljából. Beszervező akciójuk célpontjait leginkább a közösségi médián és egyéb online kommunikációs csatornákon keresztül szólítják fel terrorcselekmények végrehajtására. Csak Európában 2015 óta több mint 40 terrorcselekményt követtek el radikális nézeteket valló személyek.

A közösségimédia-szolgáltatók már régóta eltökéltek a terrorizmus elleni küzdelemben, de elismerik, hogy nincs külön eszközük az interneten található terrorcselekményre buzdító tartalmak hatékony kiszűrésére. Az Európai Unió legújabb *Horizon 2020* elnevezésű kutatásának, a RED-Alert programnak a célja a terrorista hálózatok online toborzását megakadályozó szoftver fejlesztése. A program innovatív technológiák egyedülálló kombinációja, amely kihasználja a mesterséges intelligencia erejét, és ötvözi azt a szemantikus médiaelemzés, a közösségi hálózatelemzés és a komplex eseményfeldolgozás eszközeivel. A hálózatkutatás módszertana jelentősen hozzájárul ahhoz, hogy előre tudjuk jelezni a kapcsolatok alakulását, ami hozzásegítheti a bűnüldöző szerveket ahhoz, hogy időben megállítsák az interneten szerveződő terrorcselekményeket, illetve a terrorista csoportok toborzási tevékenységét.

Biztonságpolitika: kritikus infrastruktúrák biztonsága (kiberbiztonság, országos energiaellátás)

Az országos villamosenergia-rendszerek instabillá válhatnak, ha a magasfeszültségű alaphálózat megsérül. Ilyenkor az országos villamosenergia-rendszerek úgynevezett szigetekre

szakadhatnak szét, ami veszélyeztetheti a lakosság és az ipar folyamatos energiaellátását. A villamosenergia-hálózatok tervezése és fejlesztése esetén szükséges megvizsgálni, hogy milyen hálózatfejlesztés erősíti a rendszer stabilitását. A hivatkozott elemzés a német és az olasz magasfeszültségű hálózatokat vizsgálta, és modellszerűen bemutatta, hogy ha az alaphálózat egyes elemeit véletlenszerűen vagy bizonyos tulajdonságaik alapján a rendszerből kivesszük, akkor az így születő hálózati szigetek mennyire önfenntartók.[4] A hálózatkutatási módszertan alkalmas a meglévő elektromos hálózatok fejlesztése, illetve újak kiépítése esetén a hálózatbiztonsági szempontok alapján történő modellezésre.



3. ábra

A német villamosenergetikai rendszer „hálózatos" ábrázolása

Forrás: Mureddu et al. 2016

A hazai villamosenergia-hálózat jelenleg is folyamatos fejlesztés alatt áll (lásd például az alternatív energiák kihasználására vonatkozó új beruházásokat, illetve Paks II-t).

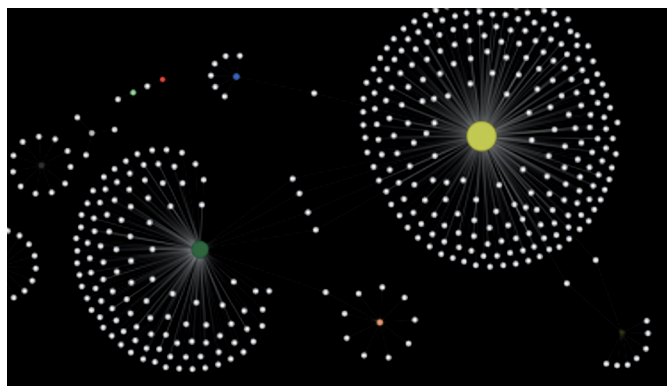
Egészségügy: egészségkommunikáció

Példák az egészségügy területén végzett hálózati kutatásokra:

- Tiltott ipari kommunikáció (például dohányipar): itt azt elemezték, hogy a dohányipar hogyan használja a közösségi média felületeit arra, hogy eljuttassa üzeneteit a legsérülékenyebb célcsoport (18 év alattiak) számára, a fiatalokhoz.
- Vakcinaellenes mozgalmak: az ebolajárvány körül kitört nemzetközi pánikreakciók során az egyik legtöbbet hallott kérdés: miért nincs erre még gyógyszer? A legtöbbet hallott követelés pedig az: mikorra lesz már védőoltás az adott betegség ellen? Bármilyen távoli is legyen az ebola, a betegség ijesztő volta és a körülötte folyó sajtópánik torzítja érzékelésünket, sokkal jobban félünk ettől a rejtélyes kórtól, mint az évente ezrek életét követelő, saját közösségünkben is gyakori megbetegedéseinktől. A hiva-

[4] Mureddu et al. 2016.

talos magyar védőoltási rendszer fejlesztése abba a fázisba érkezett, hogy elérhető a csak a méhnyakrák területén több száz életet követelő humán papillomavírus (HPV) egyes típusai elleni védőoltás. A védőoltás elleni kritikus hangok és pánikkeltési kísérletek a magyar nyelvű online térben is megjelentek, és szinte azonnal megérkeztek rá a cáfolatok, bizonyítva a hírhamisítási kísérleteket. Mivel a járványos fertőző betegségek és a HPV ellen való küzdelem a vírus elterjedtsége és lehetséges súlyos következményei miatt kulcsfontosságú, érdemes megnézni, milyen hatása lehet az ilyen hírek terjedésének a megbízhatóan működő magyar oltási rendszerre.



4. ábra

HPV-oltás-ellenes aktivitás elemzése a közösségi médiában

Megjegyzés: Az ábra azt mutatja, hogy két nagy Facebook-csoport volt elkülöníthető, amelyek ellenezték a HPV-oltást. A két nagy csoport között négy profil közvetítette a híreket. Az ábra jól ábrázolja azt is, hogy a közösségi médián keresztül milyen hatékony lehet az elérés: 2-6 felhasználó százas nagyságrendben tud híreket közvetíteni, amelyek hitelességét a személyes ismertség jelentősen befolyásolja.

Forrás: Network science: crucial to pharma. West Byfleet: pharmapforum.com, 2015. Elérhető: https://pharmaphorum.com/views-and-analysis/network-science-crucial-to-pharma (A letöltés dátuma: 2019. 03. 21.)

Egészségipar: gyógyszerfejlesztés és klinikai vizsgálatok

A hálózatok különösen fontosak a gyógyszerek fejlesztésében. A hálózati gyógyszerkutatás végső célja olyan készítmények kifejlesztése, amelyek súlyos mellékhatás nélkül gyógyítják a betegségeket.[5] A kutatás sok szinten fut, idetartozik a sejteken belüli molekuláris biológiai kölcsönhatási hálózatok dollármilliókat felemésztő feltérképezése, az eszközök és adatbázisok fejlesztése, az adatok tárolása, javítása és elemzése.

A gyógyszerinnováció eredményességének növelése érdekében különösen fontos a klinikai vizsgálatok hatékonyságának növelése, hiszen a gyógyszerinnováció körülbelül 2,5 milliárd dolláros költségének jelentős hányadát azok a klinikai vizsgálatok emésztik fel, amelyek nem csekély hányada eredménytelen (azaz a klinikai vizsgálatok alá vetett molekuláknak csupán szerény hányada válik gyógyszerré). A klinikai vizsgálatok száma és minősége

[5] HOPKINS 2007.

tekintetében hazánk mindig is az európai élvonalhoz tartozott, amiben a vizsgálóhelyek és a nemzeti hatóság felkészültsége egyaránt szerepet játszott. A hazai klinikai gyógyszervizsgálatok tradicionálisan magas számának növelése jelentős bevételi forrást jelenthet a jövőben, és nem utolsósorban a részt vevő betegeknek és intézményeknek is közvetlen és közvetett előnyöket jelenthet. Ehhez azonban szükség van arra, hogy minél szélesebb körben érvényesüljön a precíziósmedicina-szemlélet a vizsgálatok tervezése során, valamint szükség van digitális egészségügyi és életmódadatokra, ezekhez pedig elemző és prediktív adathasznosítási módszerek kellenek: a potenciálisan bevonható betegek azonosítása, azok teljes adatkészletének sok szempontú, korszerű analitikai feldolgozása, virtuális klinikai vizsgálat kivitelezése stb. Az elmúlt években nagy figyelmet kapott a mesterséges intelligencia alkalmazása a gyógyszerkutatásban, elsősorban a vizsgálattervezésben (a fejlesztési költségek óriási csökkenését remélve). Kellően nagyszámú adat alapján a mesterséges intelligencia képes előrevetíteni, hogy az adott gyógyszer az adott beteg esetében hatásos vagy hatástalan, tehát hogy a klinikai vizsgálat eredményes lesz-e. Ezáltal a vizsgálat időtartamát lerövidíti, eredményességének mértékét, azaz költséghatékonyságát pedig jelentősen javítja.[6]

Közlekedés és turizmus: közösségi közlekedés

A városi közlekedés tervezése és fejlesztése kapcsán kulcskérdés, hogy egy adott város lakói milyen kiindulási pontokról milyen érkezési pontokra kívánnak eljutni, és mindezt le kell vetíteni napszakokra, a hét egyes napjaira, évszakokra, megvizsgálni az időjárási viszonyok függvényében, és az idősoros adatállományok alapján ki kell mutatni az egyes trendeket. Ezeket az igényeket kiindulási és érkezési pontokkal meghatározható, városmérettől függően sok ezer vagy millió utazás írhatja le. A hálózattudományi módszerekkel feltérképezett és vizsgált igényeket a városi közlekedés tervezése kapcsán az alábbi szereplők használhatják:[7]

- a helyi önkormányzatok;
- városi közlekedési vállalatok, például Budapesten a BKK, amelynek *Futár* szolgáltatása kiválóan alkalmas a kutatási adatok gyűjtésére.

A hálózattudomány országos szinten is ugyanúgy használható a közlekedés fejlesztésére. Itt a kutatásban érdekelt szervezetek között meg kell említeni a következőket:

- helyi önkormányzatok;
- a települések közötti közlekedést lebonyolító vállalatok, így például a MÁV és a Volán.

A városi és még hangsúlyosabban a települések közötti közlekedésfejlesztési projektek során a hálózatkutatás kiindulási paraméterei (a kiindulási és érkezési pontokkal meghatározható utazási igények) mellett a fejlesztésben érdekelt szervezetek közötti együttműködést is lehetséges és érdemes feltérképezni.

---

[6] Lásd: Cambridge Healthtech Institute 2018.

[7] Saberi et al. 2017.

5. ábra

A budapesti közösségi közlekedés hálózatos rendszere

Megjegyzés: Az ábra egyrészt jól szemlélteti, hogy egy olyan rendkívül komplex rendszer, mint a nagyvárosi közlekedés rendszere milyen áttekinthetővé válik az adatvizualizációval, másrészt alkalmas a belső összefüggések, kapcsolódási pontok (átszállási lehetőségek) egyszerű szemléltetésére is.

Forrás: Szegedi 2016

Közlekedés és turizmus: turizmus

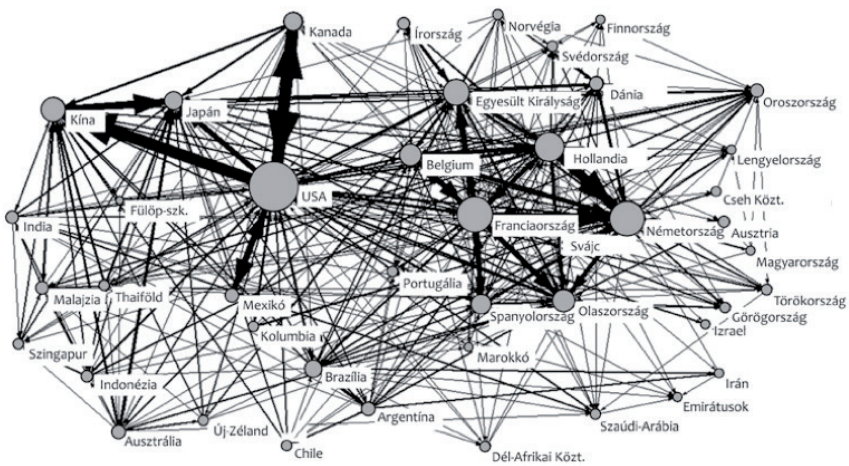
Egy adott országba vagy régióba irányuló turizmus jellemzői nagymértékben függenek a területen dolgozó szervezetek összehangolt együttműködésétől. Az együttműködés hiánya vagy a szervezeti szereplők túlzott függetlensége problémát jelenthet, hiszen visszavetheti az ágazat fejlődését, növekedését. A hivatkozott tanulmány az Elba szigetére irányuló turizmus csökkenésének okait vizsgálja, amely többek között a turizmusra specializálódott szervezetek közötti gyenge és hiányos együttműködésre vezethető vissza.[8] Elba sziget gazdasági bevételének meghatározó része a turizmusból származik, emiatt a sziget jövője erősen függ a turizmusban érdekelt szereplők közötti együttműködéstől. Ebben a modellben a turizmus desztinációs régiói mint komplex rendszerek szerepelnek, ahol a turizmusra ható szervezetek a köztük kialakult kapcsolatok alapján hálózatot alkotnak.

8 Baggio–Scott–Cooper 2010.

A hálózattudomány kvantitatív módszereit alkalmazva, egy-egy régió vagy ország turizmussal foglalkozó vagy arra hatást gyakorló szervezeteit hálózatként lehet értelmezni. A megfelelően kialakított hálózati térkép segít a turizmusban érintett szervezetek közötti kapcsolatrendszer feltérképezésében és ennek alapján célzott fejlesztésében, ami a turizmus fellendülését eredményezi.

Mezőgazdaság, élelmiszeripar, élelmiszer-biztonság

Hazánk uniós csatlakozása és a világkereskedelem további erősödése, gyorsulása miatt olykor előfordul az, hogy egy élelmiszernek minősülő termék esetében gyorsan kell kideríteni, hogy az pontosan honnan származik, vagy hogy milyen hálózatkutatás-alapú döntési támogatás kell az élelmiszerlánc-biztonsági rendszerekhez.[9]



6. ábra

Az országok élelmiszer-kereskedésének hálózatos ábrázolása

Megjegyzés: A pontok országokat reprezentálnak, nagyságuk arányos az országra jellemző éves import+export dollárban mért összértékével. Az élek hasonlóan értelmezendők. Az egyszerűség kedvéért csak azok az országok szerepelnek az ábrán, amelyek forgalma egy adott szint felett volt 2006-ban.

Forrás: Baranyi et al. 2013

Következtetések, javaslatok

A fentiek alapján célként fogalmazódik meg a nemzeti adatalapú gazdaság fejlődésének elősegítése a hálózatkutatás eredményeinek és módszereinek tudatos alkalmazásával. Ez a gazdaság, a tudomány és a közigazgatás erősödő együttműködésével jöhet létre, és mindezekhez nagymértékben hozzájárul a közszolgáltatások területének fejlesztése. Cél továbbá, hogy a magyar közigazgatásban kialakuljon egy olyan hálózatkutatási kom-

[9] Jóźwiak–Milkovics–Lakner 2016.

petencia, amely képes a legkorszerűbb tudást a mindennapokban alkalmazni és a Jó Állam célkitűzéseinek megfelelően meghatározni. Ehhez olyan nemzeti infrastruktúrákat kell kialakítani, amelyek lehetővé teszik a korszerű hálózati és adattudományi ismeretek kutatását, alkalmazását, biztosítják ezek biztonságos és védett működését, és lehetőséget teremtenek arra is, hogy a köz- és magánszereplők kölcsönösen kedvező együttműködéseket alakítsanak ki a K+F+I-tevékenységek támogatása érdekében.

Az eredményes és hatékony államigazgatás és a Jó Államban megfogalmazott célok elérése szempontjából kritikus területeken hálózattudományi megközelítésű elemzési, tervezési tevékenységbe kell fogni, és ezek alapján olyan végrehajtható projektekre van szükség, amelyek közvetlenül vagy közvetetten hozzájárulnak a tudásalapú gazdaság fejlesztéséhez és a nemzet gazdasági növekedéséhez. Mindezek létrejöttekor pedig különösen fontos, hogy az elemzések révén előállt gazdasági potenciált a nemzeti érdek szerint agilisan ki kell használni.

Felhasznált irodalom

Barabási Albert-László (2017): *A hálózatok tudománya.* Budapest, Libri.

Az Európai Bizottság sajtóközleménye: *Adatok az EU-ban: a Bizottság fokozza az egészségügyi adatok rendelkezésre állásának növelése és a megosztásuk ösztönzése érdekében tett erőfeszítéseket.* Brüsszel, 2018. április 25. Elérhető: http://europa.eu/rapid/press-release_IP-18-3364_hu.htm (A letöltés dátuma: 2019. 03. 21.)

Baggio, Rodolfo – Scott, Noel – Cooper, Chris (2010): Network Science. A Review Focused on Tourism. *Annals of Tourism Research,* Vol. 37, No. 3. 802–827. DOI: https://doi.org-/10.1016/j.annals.2010.02.008

Baranyi József – Jóźwiak Ákos – Varga László – Mézes Miklós – Beczner Judit – Farkas József (2013): A hálózatkutatás, a bioinformatika és a rendszerbiológia alkalmazási lehetőségei az élelmiszer-tudományban. *Magyar Tudomány,* 9. sz. 1094–1102. Elérhető: www.matud.iif.hu/2013/09/08.htm (A letöltés dátuma: 2018. 03. 21.)

Cambridge Healthtech Institute (2018): *Artificial Intelligence and Machine Learning in Clinical Research.* Conference, Orlando, Florida, February 12–15. Elérhető: www.SCOPEsummit.com (A letöltés dátuma: 2019. 03. 21.)

Fiatal Gasztroenterológusok Munkacsoportjának 10. Jubileumi Kongresszusa, Balatonalmádi, 2015. április 17–19., Szócska Miklós előadása.

Hopkins, Andrew L. (2007): Network Pharmacology. *Nature Biotechnology,* Vol. 25. 1110–1111. DOI: https://doi.org/10.1038/nbt1007-1110

Jóźwiak Ákos – Milkovics Mátyás – Lakner Zoltán (2016): A Network-Science Support System for Food Chain Safety. A Case from Hungarian Cattle Production. *International Food and Agribusiness Management Review,* Vol. 19, No. A. 17–42.

Mureddu, Mario – Caldarelli, Guido – Damiano, Alfonso – Scala, Antonio – Meyer-Ortmanns, Hildegard (2016): Islanding the power grid on the transmission level. Less connections for more security. *Nature.com,* Scientific Reports 6. Art. No. 34797. Elérhető: www.nature.com/articles/srep34797 (A letöltés dátuma: 2019. 03. 21.)

Ribli Dezső – Horváth Anna – Unger Zsuzsa – Pollner Péter – Csabai István (2018): Detecting and classifying lesions in mammograms with Deep Learning. *Nature.com,*

Scientific Reports 8. Art. No. 4165. Elérhető: www.nature.com/articles/s41598-018-22437-z (A letöltés dátuma: 2019. 03. 21.)

Saberi, Meead – Mahmassani, Hani S. – Brockmann, Dirk – Hosseini, Amir (2017): A Complex Network Perspective for Characterizing Urban Travel Demand Patterns. Graph Theoretical Analysis of Large-Scale Origin-Destination Demand Networks. *Transportation,* Vol. 44, No. 6. 1383–1402. DOI: https://doi.org/10.1007/s11116-016-9706-6

Szegedi Imre (2016): Hálózatok világában élünk. *Innotéka,* 2016. 01. 29. Elérhető: www.innoteka.hu/cikk/halozatok_vilagaban_elunk.1290.html (A letöltés dátuma: 2019. 03. 21.)

Vákát oldal

Varga Melinda – Ruppert Péter – Barabási Albert-László

Komplex hálózatok szerkezetének elemzése és modellezése

Bevezetés

Ez a tanulmány a természet és a társadalom komplex hálózatainak univerzális tulajdonságait vázolja fel röviden, és gyakorlati példákon keresztül mutatja be a hálózatelemzés alapvető eszközeit. A tananyag nagyon erősen támaszkodik Barabási Albert-László *A hálózatok tudománya* című könyvére,[1] amely elsősorban a műszaki és a természettudományi szakokon egyetemi képzésben részt vevő diákoknak szánt tankönyv, ennélfogva nagy hangsúlyt fektet az elméleti háttér megalapozására és a részletes matematikai levezetésekre. A jelen jegyzet keretein belül ezzel szemben igyekszünk a matematikai levezetéseket a lehető legrövidebbre fogni és az összefüggések bemutatásánál a közérthetőségre törekedni.

Manapság a közösségi médiának és az internetnek köszönhetően mindenkinek van fogalma arról, mit jelent egy hálózat, és nem szükséges hosszan magyarázni, hogy például a Facebook vagy a tömegközlekedés miért kezelhetők hálózatként. Vannak azonban olyan rendszerek is, amelyek esetén elsőre talán nem teljesen triviális, hogy hálózatként is tekinthetünk rájuk, ilyenek például a fehérjék, amelyek úgynevezett fehérje-kölcsönhatási hálózatot hoznak létre (1. ábra), vagy a betegségek, amelyek szintén hálózatba rendezhetők[2] (2. ábra).

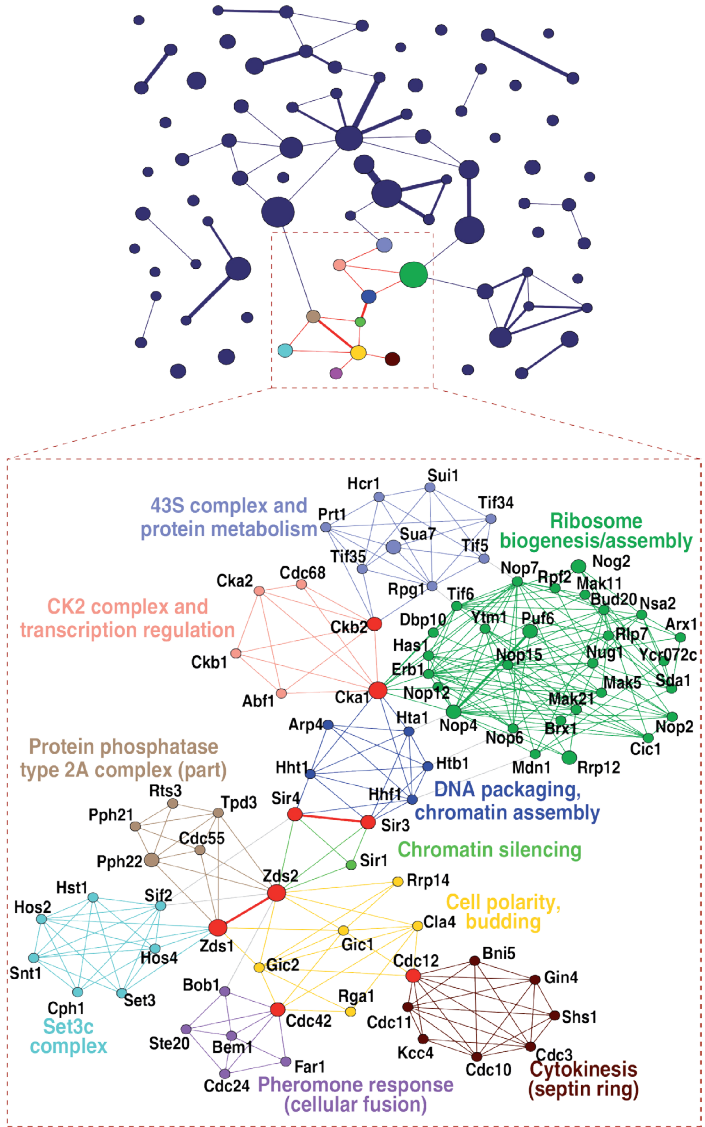
A hálózatos megközelítés segítségével tanulmányozott rendszerek sorát nagyon hosszasan lehet folytatni az ideghálózatoktól és az élő szervezetek biokémiai folyamatait leíró metabolikus hálózatoktól kezdve a tudományos társszerzőségi és egyéb kollaborációs hálózatokon, közösségimédia-hálózatokon, informatikai és kommunikációs hálózatokon keresztül a banki és céges pénzügyi tranzakciós hálózatokig, globális kereskedelmi hálózatokig és politikai hálózatokig, hogy csak a legfontosabb példákat említsük.

A hálózatos megközelítés alapgondolata az, hogy a vizsgált rendszert először alapépítőelemeire bontjuk, ezek lehetnek emberek, fehérjék, számítógépek, cégek stb., és ezeket az egységeket csomópontokkal (más néven pontokkal, csúcsokkal) reprezentáljuk. Az alapegységek közti összeköttetéseket, kölcsönhatásokat a csomópontok közé behúzott kapcsolatokkal (más néven élekkel, vonalakkal) jelöljük, így kapunk egy hálózatot (vagy matematikai nevén egy gráfot).

---

[1] Barabási 2017.

[2] Palla et al. 2005.

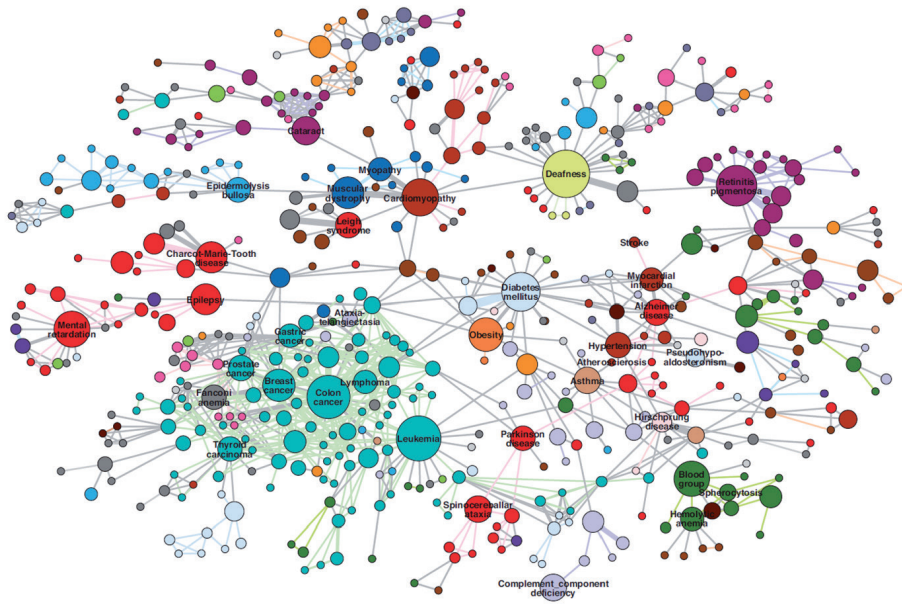


1. ábra

Egy fehérje-kölcsönhatási hálózat

Megjegyzés: A csomópontok fehérjék, és köztük a kapcsolatok azt jelzik, hogy az adott fehérjepár fizikailag össze tud kapcsolódni (ami vagy kísérleti bizonyítékokkal támasztható alá, vagy egyéb adatok alapján erősen valószínűsíthető). A fehérjék úgynevezett funkciós csoportokat alkothatnak, amelyek azonos funkciójú vagy azonos folyamatokban részt vevő fehérjékből állnak, és amelyeken belül majdnem minden csomópont majdnem minden más csomóponttal össze van kapcsolva. Ezt láthatjuk az ábra alsó részén. Az ábra felső része a funkciós csoportok közti hálózatot mutatja.

Forrás: PALLA et al. 2005



2. ábra

Részlet a betegségek hálózatából

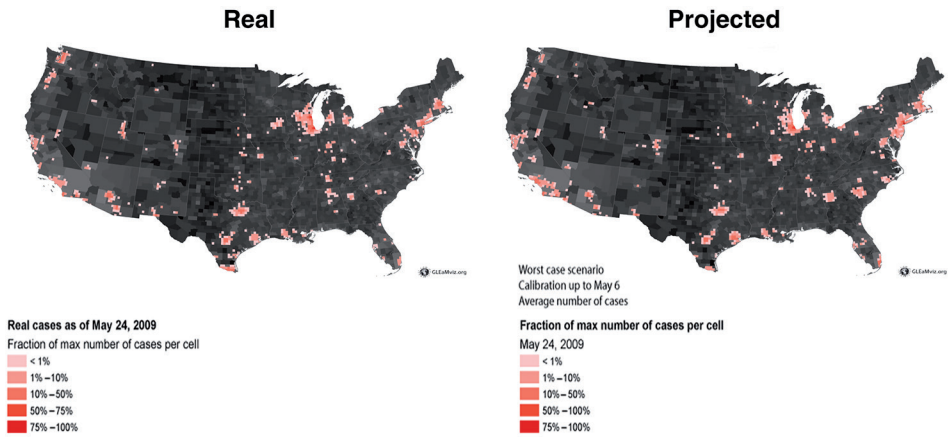
Megjegyzés: A csomópontok itt betegségeknek, illetve tüneteknek felelnek meg, a kapcsolatok pedig az elváltozások hátterében meghúzódó, közös genetikai rendellenességre utalnak. (A teljes hálózatot az 10. ábrán mutatjuk be.)

Forrás: Goh et al. 2007

A hálózatkutatás az ezredforduló óta hihetetlenül dinamikus fejlődésen ment keresztül, és manapság az egyik legfontosabb interdiszciplináris tudományterületnek számít. Ez többek között annak köszönhető, hogy a társadalom és a természet komplex rendszereit leíró hálózatok statisztikai tulajdonságai több szempontból is nagy fokú univerzalitást mutatnak. Ez annak fényében különösen meglepő, hogy maguk a tanulmányozott komplex rendszerek sokszor teljesen függetlenek egymástól, hiszen nehéz elképzelni, hogy például egy fehérje-kölcsönhatási hálózat szerkezetének bármilyen hatása lenne mondjuk a légi közlekedés hálózatára. A tapasztalt univerzalitás azt mutatja, hogy kell lennie néhány nagyon mély és általános „rendezőelvnek", amelyek ezen hálózatok szerkezetét kialakítják, és ezen elvek, tulajdonságok megismerése és megértése révén egy olyan tudáshoz jutunk, amely általánosan segíthet nagyon sok és sokféle rendszer elemzésében. A kiadványban ezen általános univerzális hálózatjellemzőket mutatjuk be, a hozzájuk kapcsolódó vizsgálati módszerekkel együtt.

A hálózatkutatás népszerűségének és sikerességének másik oka az, hogy még az alapkutatási eredményei is nagyon közel vannak a gyakorlati alkalmazásokhoz. Egy kiragadott példaként említjük itt meg Alessandro Vespignani professzor esetét, aki hálózatokon zajló

terjedési folyamatokról közölt alapkutatási eredményeket a 2000-es évek elején,[3] majd a 2009-es H1N1-járvány idején kollégáival felállítottak egy járványterjedés-előrejelző csoportot, amely napról napra követte az eseményeket és a rendelkezésre álló fertőzési adatokat, és ezek alapján adtak nagyon pontos előrejelzéseket a várható további esetek számáról (3. ábra).



3. ábra

A H1N1-járvány előrejelzése

Megjegyzés: Az ábrán a 2009-es járvány során jósolt és a tényleges esetszámok területi eloszlását láthatjuk színkód segítségével.

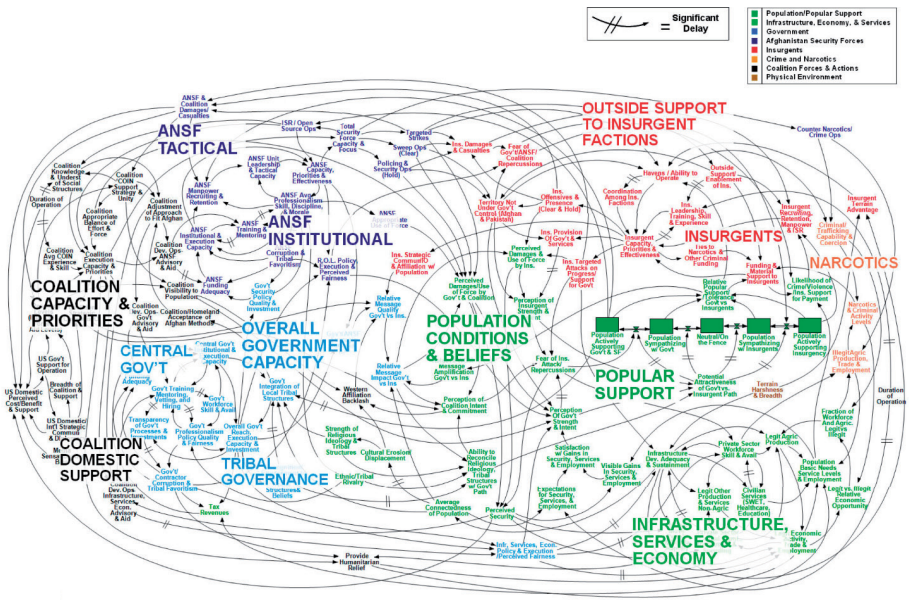
Forrás: www.gleamviz.org (A letöltés dátuma: 2019. 03. 12.)

A hálózatos gondolkodásmód alkalmazási területei közé bekerült a bűnüldöző szervek terrorizmus elleni harca is. A terrorista szervezetek pénzügyi hálói elleni küzdelem, illetve a terroristák közötti kapcsolatháló feltérképezése, leleplezése mind fontos céllá vált. Természeténél fogva ennek a területnek az eredményei nagy részben titkosítottak, de több esetben is volt már rá példa, hogy egy-egy részletesebben dokumentált esettanulmány a szélesebb nyilvánosság számára is hozzáférhetővé vált. Az egyik híres példa Szaddám Husszein elfogása,[4] akit az informális személyi kapcsolathálója alapján találtak meg, és a 2004. március 11-i madridi metrórobbantások elkövetőit is mobilhívásaik hálózatának ismeretében azonosították. A hálózatos gondolkodás a hadtudományokba is kezd beépülni.[5] Ennek egyik eredménye a hálózati alapú hadviselés, amelynek lényege, hogy decentralizált és rugalmas hálózatos szervezettel próbálják kezelni a terroristák és a bűnözői hálózatok elleni, alacsony intenzitású konfliktusokat (4. ábra).

[3] PASTOR-SATORRAS–VESPIGNANI 2001.

[4] WILSON 2010.

[5] ARQUILLA–RONFELDT 2001.



4. ábra

Egy katonai hadművelet mögötti hálózat

Megjegyzés: Ezt a diagramot az afgán háború idején, 2002-ben készítették az amerikaiak az Afganisztánban használt stratégia ábrázolására. Bár a sajtó gúnyolódott ennek az ábrának a bonyolultságán és részletességén, szemléletesen látszik rajta a modern katonai szerepvállalási feladatok szorosan összefüggő hálózata. Ma ezen a példahálózaton tanulják a tisztek és a katonai hallgatók, hogy melyek a hálózati modellek erősségei és hasznuk a döntéshozatalban, a műveletek összehangolásában. A tábornokok munkája már nem csupán a szükséges fegyveres erőről való gondoskodás tehát; figyelembe kell venniük a helyi lakosság gondolkodásmódját és életkörülményeit, valamint a lázadók katonai műveleteit finanszírozó kábítószer-kereskedelmet is.

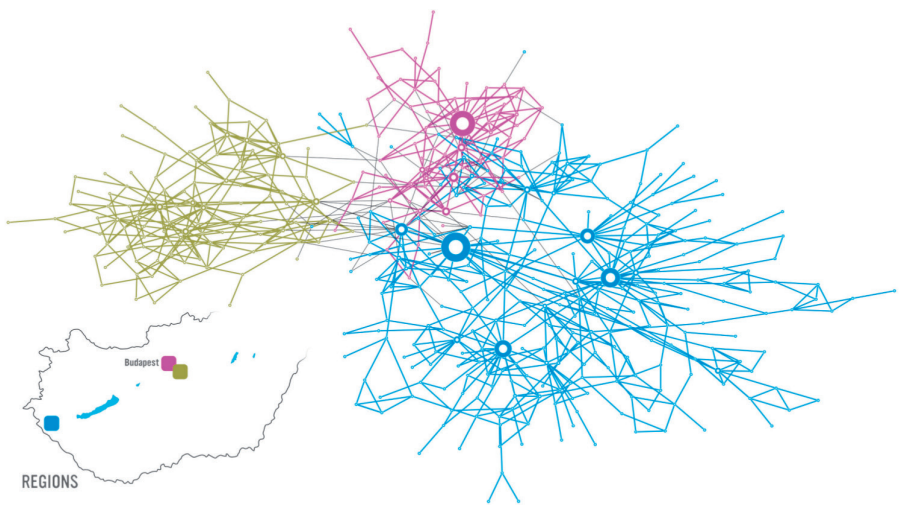
Forrás: The New York Times (A letöltés dátuma: 2019. 03. 12.)

Egy további szemléletes példa a hálózatkutatás alkalmazási területeire a szervezetfejlesztési tanácsadás. Minden szervezet vezetője hajlamos a hivatali hierarchia alapján gondolkodni, holott egyre nyilvánvalóbb, hogy egy szervezet sikerét az emberek egymás közti kommunikációja által definiált informális hálózat határozza meg. A pontos szervezetihálózat-térképek rámutathatnak arra, hogy esetleg nincs kapcsolat igen fontos egységek között, segítik azonosítani azokat, akiknek nagy szerepük van a szervezeti részlegek és a termékek összekapcsolásában, és segítik a felső vezetőket a sokféle szervezeti probléma korai felismerésében. Ráadásul a szervezetirányítás friss szakirodalma szerint egyre elfogadottabb az a nézet, hogy egy alkalmazott teljesítményét főként az informális szervezeti hálózatban betöltött szerepe határozza meg.

Több cég is (például a Maven7, az Activate Networks és az Orgnet) kínál eszközöket és módszereket a szervezetek belső szerkezetének feltérképezésére. Ezek a cégek a szolgáltatások széles skáláját kínálják: a véleményformáló egyének azonosításától kezdve az alkalmazottak felmondásának megakadályozásáig vagy a tudás és a termékek terjesz-

tésének optimalizálásáig. Javaslatot tehetnek mindenféle, a feladattól függően sokszínű, megfelelő méretű és szakértelmű csoport felállítására is. Az IBM-től az SAP-ig sok jó nevű, ismert cég egészítette ki kínálatát társadalmi kapcsolatok hálózatát elemző szolgáltatásokkal. Összességében a hálózatkutatási eszközök elengedhetetlenek a gazdaság és az irányítás területén, mert javíthatják a szervezetek termelékenységét és az innovációt.

Egy konkrét esetet láthatunk az 5. és a 6. ábrákon: egy magyar vállalat belső kommunikációs problémájának magyarázata tárul fel a hálózatkutatásnak köszönhetően. A vállalat vezetői felismerték, hogy a felső vezetők terveiből az alkalmazottakig eljutó információk gyakran lényegesen eltértek az irányítók valódi szándékaitól. Az információáramlás hatékonyságának növelésére a vállalat megkereste a szervezeti átalakításokra a hálózatkutatást felhasználó Maven7 nevű céget. A Maven7 online felületet fejlesztett ki, azon a vizsgált cég minden alkalmazottjától megkérdezték, hogy kinek a tanácsát kéri ki, ha egy a céggel kapcsolatos döntést kell meghoznia. Ez a gyűjtés szolgáltatta az 5. ábrán látható hálózatot. Ezen két személy akkor van összekapcsolva, ha legalább egyikőjük megjelölte a másikat szervezeti és szakmai információforrásként. A térkép néhány erősen befolyásos személyt mutatott ki: az ábrán ezek az erős emberek nagyobb csomópontként jelennek meg.



5. ábra

Egy magyar vállalat munkatársainak informális kapcsolathálózata

Megjegyzés: A dolgozók három fő helyszínen végezték a munkát, ezt jelöli a három szín (sárga, lila és kék). A nagyobb körökkel jelölt csomópontok kiemelkedően fontosak az információáramlás szempontjából.

Forrás: BARABÁSI 2017

Érdekes kép tárul elénk, ha a vállalat formális hierarchiáját összevetjük az informális hálózatban mérhető centralitással, amely azt ragadja meg, hogy az adott csomópont mennyire tölt be fontos, központi szerepet a hálózat topológiájában (6. ábra). A vállalat informális hálózatában a vezetőségnek kicsi a szerepe, a pontok színe itt az egyéneknek a szervezetben betöltött szerepét jelöli. Észrevehetjük, hogy a piros színnel jelölt igazgatók és a kék színnel

ábrázolt felső vezetők közül senkinek nincs központi szerepe az információs hálózatban. A középpontok mind az alsóbb szintekről kerülnek ki: középvezetők, csoportvezetők és munkatársaik. A legerősebb középpont, vagyis a legbefolyásosabb személy egy átlagos dolgozó; őt szürke csúcspont jelöli az ábra közepén.

6. ábra

A formális hierarchiában betöltött szerep és a hálózati centralitás

Megjegyzés: A színezés a formális hierarchiában betöltött szerepet mutatja, a csomópontok mérete pedig a „közelség" nevű hálózaticentralitás-mutatót jelöli, amely fordítottan arányos a csomópont többi ponttól mért átlagos távolságával. (A centralitási mutatókat részletesen is tárgyaljuk a *Centralitás* című alfejezetben).

Forrás: Barabási 2017

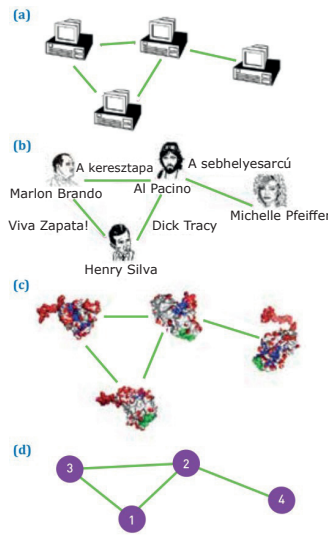
De ki ez a központi szereplő? A biztonsági és környezetvédelmi elvek betartásáért felelős ember, aki munkájából adódóan minden helyszínt rendszeresen felkeres, és beszélget a dolgozókkal. A felső vezetőkön kívül mindenkivel kapcsolatban áll. Mivel keveset tud a vezetőség valódi szándékairól, ezért azt adja tovább, amit az útja során hall, így a gyakorlatban ő a jó értelemben vett cégszintű „pletykafészek".

Hálózatok és gráfok

Ha meg akarunk érteni egy komplex rendszert, akkor ismernünk kell a részei (alkotóelemei) közötti kapcsolatokat, más szóval a rendszer kapcsolási rajzát. A hálózat felöleli egy rendszer alkotóelemeit (azokat általában pontoknak vagy csúcsoknak nevezzük), és a köztük levő kapcsolatokat (ezek az élek, angol szóval linkek). A hálózatok ezzel a módszerrel való ábrázolása közös nyelvet ad jó néhány, egymástól erősen eltérő természetű, megjelenésű vagy alkalmazhatóságú rendszer tanulmányozásához. A 7. ábrán látszik például, hogy három egészen más rendszernek is lehet ugyanolyan a hálózatos ábrázolása.

A 7. ábra két fontos hálózati jellemzőt vezet be:

- A csomópontok száma (N) a rendszer részeinek a száma; ezt a számot nemegyszer a hálózat méretének is fogjuk majd nevezni. A pontokat az $i = 1, 2, ..., N$ sorszámmal különböztetjük meg egymástól.
- A kapcsolatok (élek, linkek) számát L-lel jelöljük; ez tehát a csomópontok közötti kapcsolatok teljes száma. Az éleknek általában nincsen külön jelük, mivel az általuk összekötött csomópontokkal is megjelölhetők; például a (2, 4) él a 2. és a 4. csomópontot köti össze. A 14. ábrán bemutatott hálózatban $N = 4$ és $L = 4$.



7. ábra

Különböző rendszerek azonos hálózattal

Megjegyzés: a) Útválasztók (routerek, speciális számítógépek) kapcsolódása az internet egy kis tartományában. b) A hollywoodi színészek hálózata – abban két színész pontosan akkor van összekapcsolva, ha együtt szerepeltek legalább egy filmben. c) A fehérje-fehérje kölcsönhatási hálózat; két fehérje akkor van benne összekötve, ha kísérletek bizonyítják, hogy a sejtekben ez a két fehérje összekapcsolódik. Bár a pontok és a kapcsolatok sokfélék, a gráf mégis ugyanaz: d) $N = 4$ csomóponttal és $L = 4$ kapcsolattal.

Forrás: BARABÁSI 2017

Egy hálózat élei lehetnek irányítottak és irányítatlanok is. Bizonyos hálózatokban irányítottak az élek; a WWW-ben például az URL-ek egyik webes dokumentumról a másikra mutatnak. Irányított élei vannak a telefonhívások hálózatának is, hiszen az egyik ember hívó, a másik meg hívott fél. Más esetekben viszont irányítatlanok az élek. Például ha én járok Zsuzsival, akkor Zsuzsi is jár velem, és a villamos hálózat vezetékein is mindkét irányban folyhat az áram. Egy hálózatot akkor mondunk irányítottnak (másképpen: digráfnak), ha minden éle irányított, irányítatlannak pedig akkor, ha minden éle irányítatlan. Némely hálózatnak vannak irányított és irányítatlan élei is. Az anyagcsere-hálózatban egyik-másik folyamat megfordítható (vagyis kétirányú, irányítatlan), mások viszont nem fordíthatók meg, azaz csak az egyik irányban működhetnek (irányítottak). Egy rendszer hálózatos ábrázolásakor hozott döntések már nagyrészt meghatározzák, hogy hálózatkutatási

módszerekkel kaphatunk-e majd választ a rendszerre vonatkozó kérdésekre. Ha például emberekről van szó, akkor az, hogy milyen kapcsolatot veszünk figyelembe, eleve meg fogja határozni a megválaszolható kérdések körét.

- Ha azokat az embereket kötjük össze, akik a munkájuk révén rendszeresen kapcsolatba kerülnek egymással, akkor szervezeti vagy szakmai hálózatot kapunk; ez kulcsfontosságú lehet a vállalatok és egyéb szervezetek sikerében, és fontos terület a szervezetfejlesztési kutatásokban.
- Ha a barátokat kötjük össze, akkor a barátságok hálózatát kapjuk; az fontos szerepet kaphat az ötletek, termékek és szokások terjedésének vizsgálatában, és nagyon érdekes a szociológusok, marketingesek és az egészségügy szempontjából.
- Az intim kapcsolatban állók összekötésével szexhálózatot kapunk. Ez a hálózat nagyban meghatározza a szexuális úton továbbadható betegségek, például az AIDS terjedését; ezért is vizsgálják nagy erővel a járványkutatók.
- A telefonos és az e-mail-adatok alapján összeköthetők az egymásnak telefonálók vagy e-mailt küldők; ezzel az ismeretségek hálózatát kapjuk. A marketinggel és a kommunikációval foglalkozóknak célszerű jól ismerniük ezt a szakmai, baráti és intim kapcsolatokat is felölelő hálózatot.

Bár a felsorolt négy hálózatban sok él egyezhet (két munkatárs között lehet baráti, sőt intim kapcsolat is), maguk a hálózatok használatukban és céljukban is eltérnek egymástól.

Gráftípusok

Sok esetben az élekhez erősség, súly vagy egyfajta intenzitás is társítható; az ilyen rendszereket általában súlyozott gráfokkal reprezentálhatjuk. A mobilhívások hálózatában ez a súly lehet mondjuk két ember beszélgetésének a hossza, az áramellátási hálózatban pedig a vezetéken éppen átfolyó áram erőssége. Emellett az is előfordulhat, hogy két csúcs között több élt is figyelembe vehetünk párhuzamosan; például egy emberek közötti hálózat esetén külön élekkel jelölhetjük a telefonálást és az e-mail-kapcsolatot ugyanazon két személy között. Továbbá előfordulnak olyan hálózatok is, amelyeknél olyan élek jelenlétét is érdemes figyelembe venni, amelyek mindkét végükkel ugyanazon csúcshoz kapcsolódnak; ilyen lehet például egy tápláléklánc, amelyben bizonyos ragadozó fajok felnőtt egyedei megehetik fiatal fajtársaikat.

A fentiek alapján látható, hogy egy komplex rendszer tanulmányozása során általánosságban igen gazdag a választék különféle gráftípusok között, amelyek szóba jöhetnek potenciális „kapcsolási rajzként”. Annak érdekében, hogy egy könnyen áttekinthető listánk legyen a hálózatelemzés során lehetséges gráfok különböző szempontok szerinti osztályozásáról, megadunk egy táblázatszerű felsorolást is a vázolt esetekről.

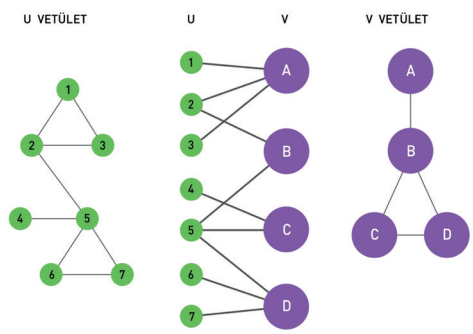
1. táblázat
Gráftípusok

Gráftípus	**Leírás**
irányított gráf	az élek irányítottak, mert a kapcsolat (valamilyen szempont szerint) aszimmetrikus
irányítatlan gráf	az élek irányítatlanok (szimmetrikus kapcsolatok)
kevert gráf	irányított és irányítatlan élek is előfordulnak
súlyozott gráf	az élekhez egy súlyérték is társul, amely a kapcsolat erősségét jellemzi
súlyozatlan gráf	az éleknek nincs súlya (szokás bináris gráfnak is hívni)
egyszerű gráf	maximum 1 élt engedünk meg ugyanazon csúcspár között (irányított esetben irányonként)
multigráf	a többszörös élek is megengedettek
pszeudográf	a többszörös élek mellett az önkapcsolatok is megengedettek

Forrás: a szerző szerkesztése

Páros gráfok

Külön szót kell ejtenünk a páros gráfokról (más néven bipartit gráfokról), amelyek olyan hálózatnak felelnek meg, amelyben a csomópontok feloszthatók két különálló *U* és *V* halmazra úgy, hogy a hálózat minden éle egy *U*-beli pontot köt össze egy *V*-beli ponttal. Másként fogalmazva: ha az *U* halmaz pontjait zöldre színezzük, a *V* halmaz pontjait pedig lilára, akkor a hálózat minden éle egy zöld és egy lila színű pontot köt össze (8. ábra).

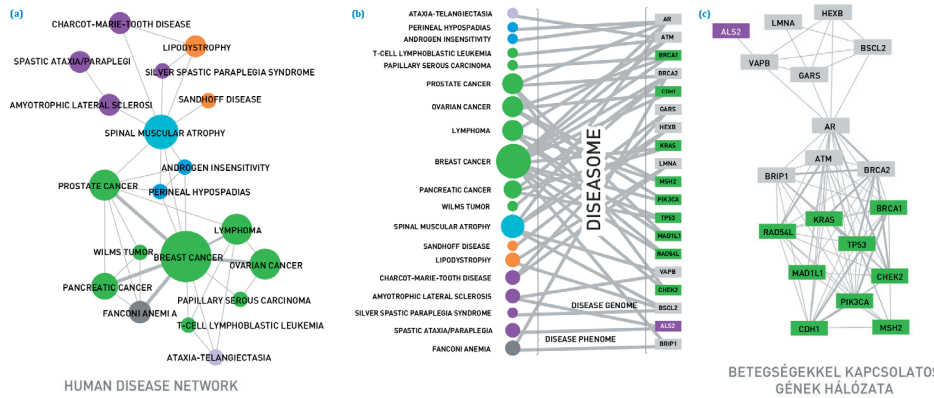


8. ábra
Páros (bipartit) hálózat

Megjegyzés: Egy páros hálózat pontjai két halmazra oszthatók (*U*-ra és *V*-re). Az *U* halmazbeli csomópontok csak a *V* halmazbeliekkel vannak közvetlen kapcsolatban, vagyis nincsenek közvetlen *U* – – *U* és *V* – – *V* kapcsolatok. Az ábra a páros hálózatokból készíthető két vetületet mutatja. Az *U* vetületben akkor kötjük össze az *U* halmaz két pontját, ha a páros ábrázolásban van egyező *V*-beli szomszédjuk; a *V* vetületben meg akkor kapcsoljuk össze a *V* halmaz két pontját, ha a páros hálózatban van egyező *U*-beli szomszédjuk.

Forrás: BARABÁSI 2017

Minden páros hálózathoz készíthető két vetület. Az első vetület *U*-nak azokat a pontjait köti össze egymással, amelyeknek van közös *V*-beli szomszédjuk a páros gráfban. A másik vetület meg *V*-nek azokat a pontjait köti össze, amelyeknek van közös *U*-beli szomszédjuk a páros gráfban (lásd a 8. ábrát).



9. ábra

Az emberi betegségek hálózatának egy rákkal kapcsolatos részlete

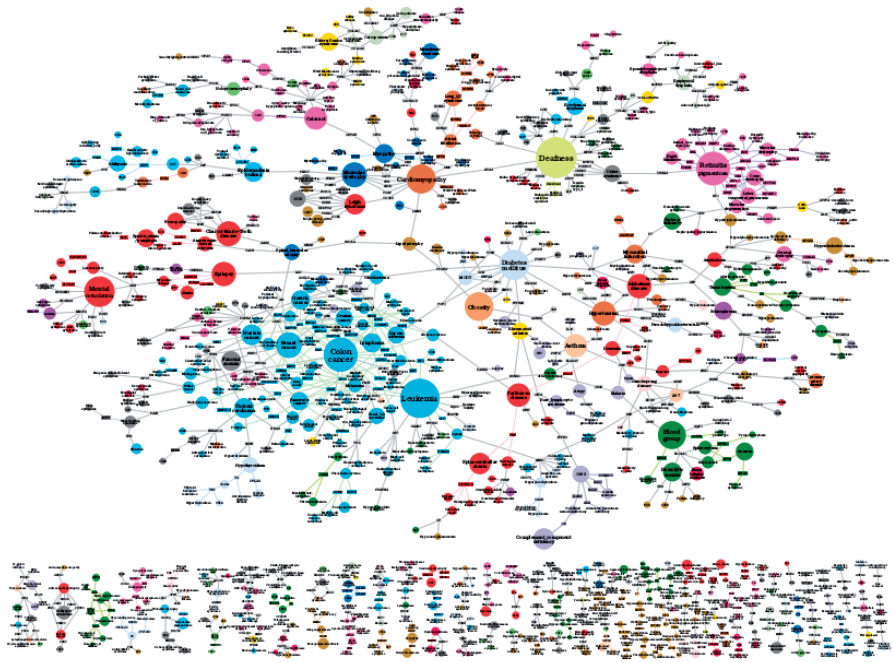
Megjegyzés: a) A bipartit betegséggéntérképből készíthető egyik vetület a betegségek hálózata, amelyben két betegség között akkor van kapcsolat, ha ugyanahhoz a génhez köthetők, azaz hasonló a genetikai eredetük. b) A betegséggéntérkép, más néven diseasome (a disease és genome szavak összevonásából) egy páros hálózat, pontjai a betegségek és a gének. Egy betegség akkor kapcsolódik egy génhez, ha a gén mutációinak köze lehet a szóban forgó betegséghez. c) A másik vetület a gének hálózata. Két gén között akkor van él, ha közös betegségekhez kapcsolódnak.

Forrás: Goh et al. 2007

A hálózatelméletben sok páros hálózattal találkozhatunk. Általánosan ismert példa a hollywoodi színészek hálózata; abban a pontok egy része a filmeket jelöli, más részük a színészeket. Egy filmet akkor kötünk össze egy színésszel, ha a színész szerepelt abban a bizonyos filmben. Ennek a páros hálózatnak az egyik vetülete a színészek hálózata. Abban a pontok (színészek) akkor vannak egymással kapcsolatban, ha szerepeltek ugyanabban a filmben. A másik vetület a filmek hálózata; abban két film akkor van összekapcsolva, ha volt olyan színész, aki játszott mindkettőben.

Az orvostudományban is találni fontos példát páros hálózatokra. Például az emberi betegségek hálózata ezt vagy azt a betegséget azokkal a génekkel köti össze, amelyeknek a mutációjából (közvetve vagy közvetlenül) fakad,[6] ahogy azt a 9. ábra illusztrálja egy rákkal kapcsolatos alhálózaton. A teljes bipartit betegséggén-hálózat összesen 1283 elváltozást köt össze 1777 okozó génnel, aminek a betegségek halmazára vett vetületét a 10. ábra mutatja be.

[6] Goh et al. 2007.



10. ábra

A teljes emberi betegségek hálózata a betegségek halmazára vetítve

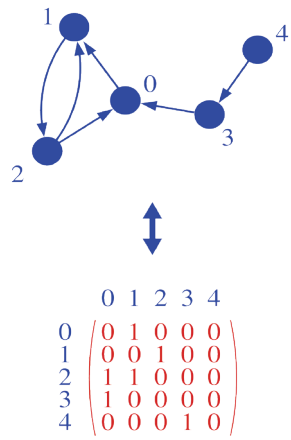
Forrás: GOH et al. 2007

Szomszédsági mátrix

Ha egy hálózatot teljeskörűen akarunk leírni, ismernünk kell a hálózatbeli kapcsolatokat. Ennek az a legegyszerűbb módja, ha teljes listát készítünk a kapcsolatokról. Például a 7. ábra hálózatának négy kapcsolata egyenként is felsorolható: {(1,2),(1,3),(2,3),(2,4)}. Matematikai megfontolásokból a hálózatokat gyakran a szomszédsági mátrixszal írjuk le. Egy N pontból álló irányított hálózat szomszédsági mátrixának N sora és N oszlopa van, és elemei a következők:

$$A_{ij} = \begin{cases} 1, & \text{ha van él } i \text{ – ből } j \text{ – be,} \\ 0, & \text{egyébként.} \end{cases} \qquad 1.$$

A fenti definícióból következik, hogy egy irányítatlan hálózat szomszédsági mátrixában minden él kétszer szerepel, és maga a mátrix szimmetrikus (azaz a főátlójára tükrözve önmagába megy át, hiszen $A_{ij} = A_{ji}$). Természetesen egy irányított hálózat szomszédsági mátrixa általános esetben már nem lesz szimmetrikus, ahogy azt láthatjuk például a 11. ábrán.



11. ábra

Egy kis méretű irányított hálózat és a neki megfelelő szomszédsági mátrix

Forrás: a szerző szerkesztése

Amennyiben súlyozott hálózatot elemzünk, a szomszédsági mátrix elemei a súlyozatlan hálózatoknál látott bináris 0 vagy 1 értékek helyett a megfelelő kapcsolat súlyát adják meg, azaz

$$A_{ij} = w_{ij}, \tag{2.}$$

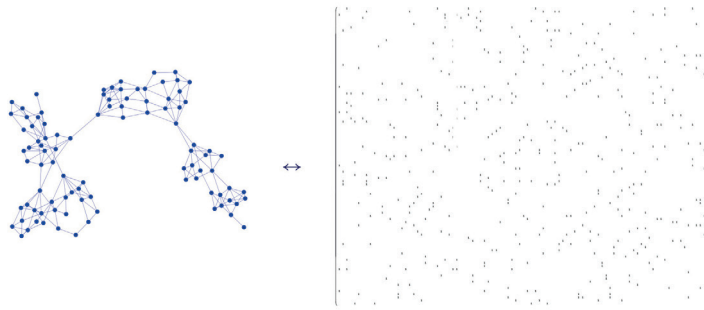
ahol w_{ij} az i-ből induló és j-be érkező él súlya. (Természetesen, ha i és j között nincs él, akkor a mátrixelem továbbra is 0 értéket vesz fel.)

Ritka és sűrű gráfok

Egy valóságos hálózatban a csomópontok N és a kapcsolatok L száma igen széles határok közé eshet. Például a *C. elegans* féreg idegsejtjeinek hálózata az egyetlen olyan hálózat, amely teljesen feltérképezi egy élőlény idegrendszerét. Ennek a hálózatnak $N = 302$ neuronja (pontja) van. Az emberi agy sokkalta nagyobb, nagyjából százmilliárd ($N = 1011$) neuronból áll (és még nincsen feltérképezve). Egy emberi sejt genetikai hálózatát 20 ezer gén (pont) alkotja, ismeretségeink hálózata hétmilliárd emberből áll ($N \simeq 7{\cdot}109$), a WWW pedig egybilliónál is több webdokumentumot tartalmaz ($N > 1012$).

Vizsgáljuk meg ezek után, hogy mi a helyzet az élek számával, valamint az élsűrűséggel. A 12. ábrán egy kis méretű fehérje-kölcsönhatási hálózatot láthatunk a megfelelő szomszédsági mátrixszal együtt.

Szembetűnő, hogy a mátrixelemek túlnyomó többsége 0, és csak elvétve találunk olyat, amely a neki megfelelő sor- és oszlopindexhez tartozó csúcsok között egy kapcsolatot jelez. Az ilyen jellegű mátrixokat *ritka mátrixoknak* hívjuk, a ritka szomszédsági mátrixszal rendelkező hálózatokat pedig *ritka hálózatoknak*. Fő jellemzőjük, hogy a csúcsok között behúzható összes jellemző éleknek csak nagyon kis hányada létezik ténylegesen is a hálózatban.



12. ábra

Egy kis méretű irányított hálózat és a megfelelő szomszédsági mátrix

Megjegyzés: A mátrix 0 elemeit fehér, a nem nulla (A_{ij}=1) elemeket pedig fekete szín jelöli.

Forrás: a szerző szerkesztése

Képzeljük most el, hogy a 12. ábrán látható mátrixban a 0 elemeket 1-re, az 1 elemeket pedig 0-ra cseréljük. Így felcserélődnek a szerepek, és egy olyan mátrixot kapunk, amelyben már csak elenyésző a 0 elemek hányada; az ilyen mátrixokat a matematikában *sűrű mátrixnak* szokás hívni. Ennek megfelelően a sűrű szomszédsági mátrixszal rendelkező gráfokat *sűrű gráfnak* tekintjük, amelyek közös jellemzője, hogy az élek száma közelít a maximálisan lehetséges $L_{\max}$ értékhez. Az $L_{\max}$ az összes lehetséges (különböző) csúcspár számával egyezik meg, ami irányított esetben egyszerűen $L_{\max} = N(N-1)$, míg irányított hálózatoknál $L_{\max} = N(N-1)/2$ (hiszen ott a páron belüli sorrend már nem számít).

Vegyük észre, hogy akár egy sűrű irányítatlan, akár egy sűrű irányított hálózattal van dolgunk, $L_{\max}$ mindkét esetben gyakorlatilag N^2-tel arányos, ami egy növekvő hálózat esetén sokkal gyorsabban nő, mint maga az *N*. Az, hogy egy hálózat mérete időben változik, sok esetben előfordul, gondoljunk például a Facebook-kapcsolatok gráfjára, amely egy egyetemi közösségimédia-felületből nőtt ki az évek során egy globális kapcsolatrendszerré. Ha az élek és csomópontok számának növekedését hasonlítjuk össze, egy ritka hálózatban azt tapasztaljuk, hogy a ténylegesen megvalósuló kapcsolatok száma arányos marad a pontok számával, azaz $L \sim N$, annak ellenére, hogy a lehetséges élek száma ennél gyorsabban, négyzetesen nő. Ezzel szemben egy sűrű hálózatban a létező élek száma a csúcsok helyett $L_{\max}$-szal arányos, azaz $L \sim N^2$.

Vizsgáljuk meg, milyen következményekkel jár ez az eltérés az egy ponthoz tartozó kapcsolatok átlagos számára nézve. Az egyszerűség kedvéért irányítatlan hálózatot feltételezve, és figyelembe véve, hogy minden élnek végpontja van, az egy csúcsra jutó kapcsolatok számának átlagos értéke a

$$\langle k \rangle = \frac{2L}{N} \qquad 3.$$

alakban adható meg. Ebből látható, hogy

- egy növekvő sűrű hálózat esetén hosszú távon $L \sim N^2$, aminek révén $\langle k \rangle = \frac{2M}{N} \sim N$, azaz az egy pontra jutó kapcsolatok átlagos száma is nagyjából *N*-nel arányos,

- míg egy növekvő ritka hálózat esetén hosszú távon $L \sim N$, aminél fogva $\langle k \rangle = \frac{2M}{N} \sim$ konstans, azaz az egy pontra jutó kapcsolatok száma nagyjából méretfüggetlen.

Végezetül bevezethetjük az *élsűrűség* fogalmát, amely a hálózat éleinek száma osztva a maximálisan lehetséges élek számával:

$$p = \frac{L}{L_{\max}} = \begin{cases} \frac{2L}{N(N-1)} & \text{irányítatlan hálózatban} \\ \frac{L}{N(N-1)} & \text{irányított hálózatban.} \end{cases} \quad 4.$$

A fenti definícióból következően p mindig 0 és 1 közötti értéket vesz fel, és megadja két véletlenszerűen választott pont között a közvetlen kapcsolat létezésének valószínűségét.[7]

A valós hálózatok ritkák

Térjünk rá ezek után a természet és a társadalom komplex rendszereit jellemző hálózatok sűrűségének vizsgálatára. Joggal merülhet fel a kérdés, hogy amennyiben a rendelkezésre álló adatok nem teszik lehetővé a csúcs- és élszámváltozás nyomon követését (mert például maga a hálózat nem egy növekvő hálózat, vagy összesen csak egy időpontban történt adatfelvétel a hálózatról), akkor miként tudjuk eldönteni róla, hogy ritka-e, vagy sűrű? A gyakorlati esetek döntő többségében általában ilyenkor is nagy biztonsággal megállapítható, hogy a vizsgált hálózat ritka, ugyanis a valódi komplex rendszereket leíró hálózatok esetén az egy pontra jutó kapcsolatok átlagos száma szinte mindig több nagyságrenddel kisebb, mint N.

Ezt a jelenséget illusztrálja a 2. táblázat, amelyben tíz különböző komplex hálózat alapadatait gyűjtöttük össze. A felsorolt mintahálózatok (referenciahálózatok) magukban foglalnak társadalmi rendszereket (mobilhívások gráfja és e-mail-hálózat), együttműködési és a részvételen alapuló hálózatokat (tudományos együttműködések hálózata, a hollywoodi színészek hálózata), információs rendszereket (WWW), technológiai és infrastrukturális rendszereket (internet, áramellátási hálózat), biológiai rendszereket (fehérje-kölcsönhatások, anyagcsere) és a tudományos cikkek egymásra való hivatkozásainak hálózatát. Ezek a hálózatok igen különböző méretűek: az E. coli baktérium anyagcsere-hálózata csupán $N = 1039$ pontból áll, a tudományos hivatkozások hálózata viszont csaknem félmillióból. Ez a tíz hálózat sok olyan kutatási területet lefed, ahol rendszeresen alkalmazzák a hálózatokat. A kutatók gyakran ezen hálózatok tulajdonságaihoz hasonlítják az általuk éppen vizsgált rendszer fontos jellemzőit. A 2. táblázatban látható, hogy a tíz referenciahálózat között akad irányított és irányítatlan is. A táblázat utolsó oszlopában látható $\langle k \rangle$ értékeket összehasonlítva az 5. oszlopban mutatott N értékekkel megállapítható, hogy ezek a hálózatok mind ritkák, hiszen $\langle k \rangle$ több nagyságrenddel kisebb, mint N.

[7] Természetesen, ha konkrétan kiválasztunk két pontot, akkor azok vagy össze vannak kötve, vagy nem; ellenben ha ezt sokszor megismételjük egymás után, akkor azon esetek aránya, amikor találtunk közvetlen kapcsolatot (kellően hosszú próbálkozássorozat esetén), tart a p-hez.

2. táblázat
A mintaként használt tíz hálózat

Hálózat neve	**Pont**	**Kapcsolat**	**Irányított/ irányítatlan**	N	L	$\langle k \rangle$
Internet	router	internetkapcsolat	irányítatlan	192 244	609 066	6,34
WWW	weboldal	hivatkozás	irányított	325 729	1 497 134	4.60
Áramellátás	erőmű, transzformátor	vezeték	irányítatlan	4 941	6 594	2,67
Mobilhívások	előfizető	hívás	irányított	36 595	91 826	2,51
E-mail	e-mail-cím	e-mail	irányított	57 194	103 731	1,81
Tudományos együttműködések	kutató	társszerzőség (közös publikáció)	irányítatlan	23 133	93 439	8,08
Színészek hálózata	színész	közös film	irányítatlan	702 388	29 397 908	83,71
Hivatkozási hálózat	tudományos cikk	hivatkozás	irányított	449 673	4 689 479	10,43
E. coli anyagcseréje	anyagcsere-molekula	kémiai reakciók	irányított	1 039	5 802	5,58
Élesztő fehérjéinek kölcsönhatásai	fehérje	kapcsolódási kölcsönhatások	irányítatlan	2 018	2 930	2,90

Megjegyzés: A jegyzet az itt felsorolt hálózatok segítségével szemlélteti a tárgyalt főbb hálózatjellemzők viselkedését. Itt most felsoroljuk a pontok és a köztük lévő kapcsolatok jelentését, az élek irányított vagy irányítatlan voltát, a pontok (N) és a kapcsolatok (L) számát, valamint az egy pontra jutó kapcsolatok átlagos számát ($\langle k \rangle$).

Forrás: BARABÁSI 2017

A ritkaság következményei

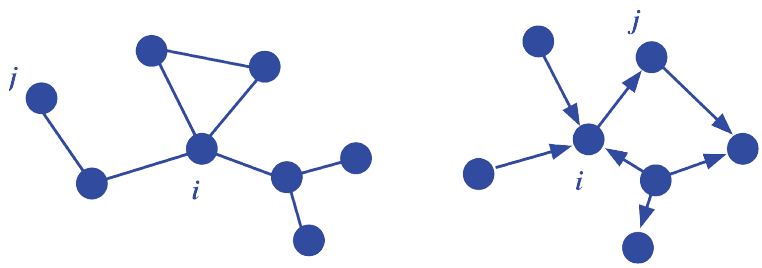
A ritkaság erősen befolyásolja a valóságos hálózatok tárolásának módját, és azt is, hogyan keresünk bennük. Például ha egy nagy hálózatot tárolunk számítógépen, akkor sokkal célszerűbb csak a kapcsolatok listáját tárolni (a nullától különböző A_{ij} elemeket), hiszen a teljes szomszédsági mátrixot fölösleges volna, lévén az A_{ij} elemeinek túlnyomó része nulla. A mátrixalakban való tárolás tehát igen nagy helyet foglal el, pedig legnagyobbrészt csak 0-kból áll. Az egyik legelterjedtebb és legkönnyebben kezelhető tárolási mód az, amikor az éleket egy két oszlopból álló file-ba mentjük el súlyozatlan esetben (az első oszlop általában a kiinduló pont, a második pedig a célpont), illetve súlyozott hálózatok esetén ezt kiegészítjük még egy 3. oszloppal, amely az élsúlyokat sorolja fel.

Egy elméleti szempontból megjegyzendő további következménye a ritkaságnak az, hogy két véletlenszerűen választott pont között a közvetlen kapcsolat valószínűsége elenyészően kicsi, és ez a valószínűség nagyjából a csúcsok számának reciprokával arányos (azaz, ha például minden határon túl növeljük a rendszerméretet, akkor $1/N$ szerint tart a nullához). Ezt talán úgy a legkönnyebb belátni, hogy annak esélye, hogy egy véletlenszerűen választott j csúcs közvetlen éllel kapcsolódjon egy adott i csúcshoz, szimplán az i-re jutó kapcsolatok száma osztva N-nel. Mivel egy ritka hálózatban az egy csúcsra jutó kapcsolatok átlagos száma nagyjából konstans (azaz független N-től, ha változik a rendszerméret), ez a hányados nagyjából $1/N$-nel lesz arányos.

Alapvető hálózatjellemzők

Fokszám

Egy hálózatban a pont legfontosabb tulajdonsága a fokszám: a pont és a hálózat többi csúcsa közötti kapcsolatok száma. Ezt az *i*-edik csúcs esetén általában k_i-vel jelöljük, és irányított hálózatok esetén természetesen különbséget teszünk a kimenő élek száma (ki-fokszám, k_i^{ki} és a bejövő élek száma (be-fokszám, k_i^{be}) között (lásd a 13. ábrát).



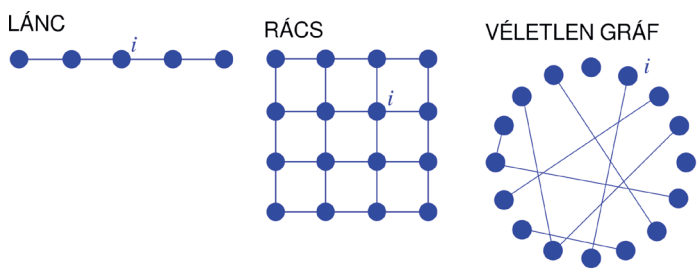
13. ábra

A fokszám fogalmának szemléltetése

Megjegyzés: A bal oldali, irányítatlan hálózatban az *i* csúcs fokszáma $k_i = 4$, míg a j esetén $k_j = 1$. A jobb oldalon látható irányított gráfban az *i* csúcs ki-fokszáma $k_i^{\text{ki}} = 1$, a be-fokszáma $k_i^{\text{be}} = 3$, illetve a *j* csúcs esetén $k_i^{\text{ki}} = k_i^{\text{be}} = 1$.

Forrás: a szerző szerkesztése

A mobilhívások hálózatában például a fokszám a szóban forgó ember híváslistájának hossza (azoknak az egymástól különböző személyeknek a száma, akikkel beszélt). Az idézettségi hálózatban pedig a be-fokszám az adott cikkre mutató hivatkozások száma, a ki-fokszám pedig azoknak a cikkeknek a száma, amelyekre a szóban forgó cikk hivatkozik.



14. ábra

A fokszám meghatározása három egyszerű esetben

Megjegyzés: Az ábra bal oldali részén látható lánc, illetve a középen mutatott rács esetén triviális, hogy a vég- és határpontokat nem számítva a csúcsok fokszáma $k = 2$, illetve $k = 4$. A jobb oldalon szereplő véletlen gráfban már csak az átlagos fokszámot könnyű megadni a $\langle k \rangle = (N-1)p$ alakban, ahol p az élbekötési valószínűség.

Forrás: a szerző szerkesztése

Bizonyos egyszerű esetekben nagyon könnyű megadni a csúcsok fokszámát (vagy legalább az átlagos fokszámot) pusztán a hálózat szerkezete alapján. Nézzünk erre három példát, amelyeket 14. ábra szemléltet:

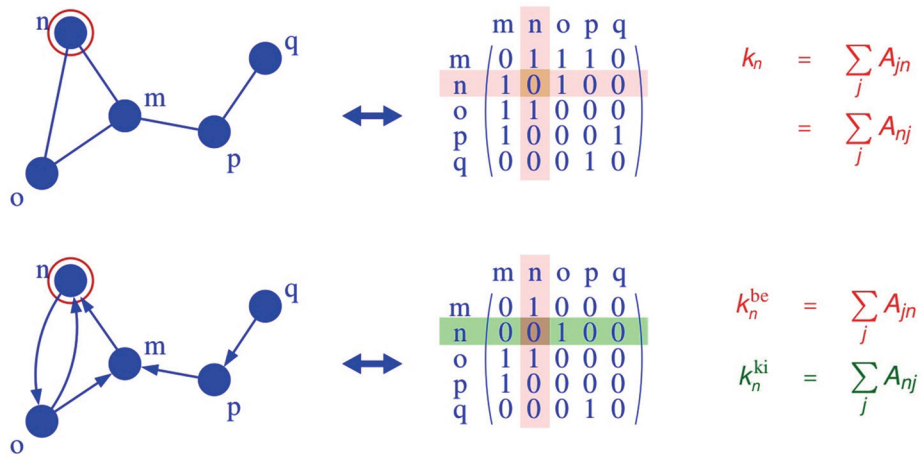
- Egy *lánc* esetén (a végpontokat leszámítva) minden csúcs fokszáma $k = 2$, hiszen mindenki a közvetlen jobb és bal oldali szomszédjához kapcsolódik.
- Egy *négyzetrács* esetén (a határon lévő pontokat leszámítva) minden csúcs fokszáma $k = 4$, hiszen itt minden csúcsnak 4 közvetlen szomszédja van.
- Egy *véletlen gráf* esetén, amely N pontot köt össze teljesen véletlenszerűen úgy, hogy minden lehetséges pár között azonos p valószínűséggel húzunk be egy élt, a pontok átlagos fokszámát tudjuk könnyen megadni a $\langle k \rangle = (N - 1)p$ alakban. Ez abból következik, hogy bármely pont $N - 1$ másik ponthoz kapcsolódhatna összesen, és ezen lehetőségeknek a p hányada valósul meg átlagosan.

Az imént vázolt véletlengráf-konstrukciót Erdős Pál és Rényi Alfréd vezette be.[8] A róluk elnevezett modellt részletesen is tárgyaljuk később az *Erdős–Rényi-modell* című fejezetben. Egyszerűsége miatt ez egy kiváló eszköz a hálózatjellemzők szemléltetésére, ennélfogva az anyagban haladva a tárgyalt fogalmak, mennyiségek viselkedését rendre meg fogjuk vizsgálni Erdős–Rényi-féle véletlen gráfok[9] esetén is, már a modell tulajdonságainak részletes ismertetése előtt.

A fenti rövid kitérő után folytassuk a fokszám fogalmával való ismerkedést a fokszám és a szomszédsági mátrix közötti kapcsolat vizsgálatával. Mivel a szomszédsági mátrix minden kapcsolatot eltárol, természetesen bármely pont fokszáma kinyerhető belőle: irányítatlan esetben a megfelelő sor vagy oszlop elemeinek összeadásával, irányított esetben pedig a ki-fokszám a sorösszegnek, a be-fokszám pedig az oszlopösszegnek felel meg, amint azt a 15. ábra szemlélteti.

[8] ERDŐS–RÉNYI 1959; ERDŐS–RÉNYI 1960.

[9] Az Erdős–Rényi-féle véletlen gráfot szokás *klasszikus* véletlen gráfnak is hívni. Erdős és Rényi eredeti cikke 1959-ben jelent meg, amelynek hivatkozottsága jelen napjainkban is intenzíven nő. A továbbiak megértéséhez egyelőre bőven elég, ha ezzel a modellel kapcsolatban annyit jegyzünk meg, hogy N csúcs között definiál egy véletlenszerű gráfot, amelyben minden lehetséges él egymástól függetlenül p valószínűséggel jelenik meg. (ERDŐS–RÉNYI 1959)



15. ábra

A fokszám és a szomszédsági mátrix kapcsolata

Megjegyzés: Irányítatlan esetben (felső sor) egy csúcs fokszámát megkaphatjuk úgy, hogy a mátrix neki megfelelő sorában vagy oszlopában összegezzük az elemeket, az itt látható gráfban például $k_n = \sum_j A_{nj} = \sum_j A_{jn}$. Irányított esetben (alsó sor) a mátrixban a sor szerinti összegzés a ki-fokszámot adja meg, $k_n^{\text{ki}} = \sum_j A_{nj}$, míg az oszlop elemeinek összege a be-fokszámnak felel meg, $k_n^{\text{be}} = \sum_j A_{jn}$.

Forrás: a szerző szerkesztése

Átlagos fokszám

Egy irányítatlan hálózatban a fokszámok átlaga megegyezik az egy pontra jutó kapcsolatok átlagos számával, amelyet a ritka és sűrű gráfokat tárgyaló alfejezetben vezettünk be a 3. számú egyenletben. Érdemes most újra felírni, és egyúttal összekapcsolni a pontonként mérhető „egyéni" fokszámokra történő átlagolással:

$$\langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_1 = \frac{2L}{N} \qquad 5.$$

Amint azt a felidézett alfejezetben részletesen is kifejtettük, az átlagos fokszám és a csúcsok számának összevetése nagyon hasznos információkat nyújt a hálózat sűrűségével kapcsolatban.

Ha a hálózat irányított, akkor az egyes csúcsok szintjén külön kell kezelnünk a ki-fokszámot és a be-fokszámot, ahogy már a fejezet elején említettük. Azonban ha az átlagos ki- és be-fokszámot vizsgáljuk, akkor abból következően, hogy minden élnek egy kezdő- és egy végpontja van valahol a csúcsok között (azaz összesen ugyanannyi kezdő- és végpont szerepel a hálózatban), arra jutunk, hogy a ki- és be-fokszámok átlagának meg kell egyeznie

egymással. Ebből fakadóan irányított hálózatok esetén a következő egyenletet írhatjuk fel az átlagos fokszámra:

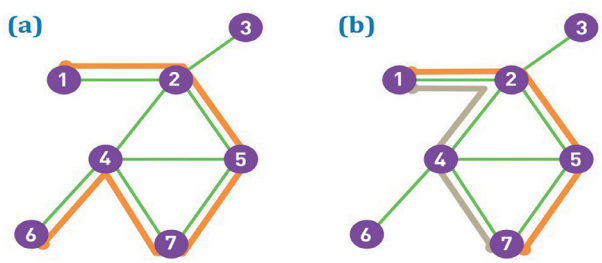
$$\left\langle k^{\text{be}} \right\rangle = \frac{1}{N}\sum_{i=1}^{N} k_i^{\text{be}} = \frac{L}{N} = \frac{1}{N}\sum_{i=1}^{N} k_i^{\text{ki}} = \left\langle k^{\text{ki}} \right\rangle \qquad 6.$$

Távolság

Az előző alfejezet, ha úgy tetszik, a csúcsok közvetlen szomszédságával foglalkozott, hiszen a közvetlen (más néven első) szomszédok száma megegyezik a fokszámmal. Lépjünk most tovább gondolatban a másod-, harmad- és általánosan az *n*-edik szomszédok felé,[10] és vizsgáljuk meg, mit tudunk mondani a csúcsok *távolságáról* a hálózatban.

A távolság fogalmának fontos szerepe van például a fizikai rendszerek elemeinek kapcsolatában, ahol természetesen a „távolság" alatt fizikai távolságot értünk. Például egy kristályban két atom vagy a világegyetemben két galaxis távolsága már meghatározza, hogy azok mekkora erővel hatnak egymásra. A hálózatokban a távolság fogalma már nem feltétlen ilyen egyszerű. Mert csakugyan, hogyan értelmezhető két weboldal távolsága vagy két, egymást nem is ismerő ember távolsága?

Ezekben az esetekben nem a fizikai távolság számít, hiszen két weboldal között akkor is lehet kapcsolat, ha a Föld két átellenes pontján vannak. Másfelől meg két ember még akkor sem feltétlenül ismeri egymást, ha mindketten ugyanabban a házban laknak. A hálózatokban a fizikai távolságot az *út hosszával* helyettesítjük. Az út a hálózat élei révén egymás után kapcsolódó, egymástól különböző hálózati pontokból áll. Az út hossza az utat alkotó élek (kapcsolatok) száma lesz (16. ábra).



16. ábra

Utak a hálózatban

Megjegyzés: a) Az ábrán narancsszínnel jelölt út az 1→ 2→ 5→ 7→ 4→ 6 pontokat köti össze, tehát a hossza: $n = 5$. b) Az 1-es és a 7-es pont közötti legrövidebb úthossz ($d_{1,7}$, a két pont távolsága) az 1-es és a 7-es pontot a legkevesebb kapcsolaton át összekötő út (vagy utak) hossza. Több legrövidebb út is lehet, az ábrán narancs- és szürke szín mutatja a két legrövidebbet. A hálózat átmérője a hálózatban található legnagyobb pont–pont távolság, itt éppen $d_{\max} = 3$.

Forrás: BARABÁSI 2017

[10] Egy *i* csúcs másodszomszédai azok a pontok, amelyek nem közvetlenül kapcsolódnak *i*-hez, de van *i* -vel közös szomszédjuk, azaz *i*-ből két lépés alatt elérhetők. Hasonlóan: *i* harmadszomszédai azok a pontok, amelyekhez *i*-ből három kapcsolaton keresztül el tudunk jutni, de kevesebb lépésben nem; stb.

Az út fontos szerepet játszik a hálózatkutatásban. A következőkben az úttal és távolsággal kapcsolatos legfontosabb fogalmakat vezetjük be:

- *Legrövidebb út:* a hálózat i-edik és j-edik pontja között a legrövidebb út a legkevesebb pontot tartalmazó út [16. ábra b) képe]. Ezt a legrövidebb utat az i-edik és a j-edik pont távolságának hívjuk, és d_{ij}-vel jelöljük. Két pont között több (azonosan d_{ij} hosszúságú) legrövidebb út is létezhet,[11] ahogy azt a 16. ábra b) képe szemlélteti. A legrövidebb út sosem tartalmaz hurkot (egyik pontja sem csatlakozik önmagához), és nem metszi önmagát (az útban minden pont csak egyszer van jelen). Irányítatlan hálózatban $d_{ij} = d_{ji}$, tehát az i-edik és a j-edik pont távolsága ugyanakkora, mint a j-edik és az i-edik pont távolsága. Irányított hálózatokban gyakran $d_{ij} \neq d_{ji}$, és az $i \to j$ út létezéséből még egyáltalán nem következik, hogy van $j \to i$ út is. A valóságos hálózatokban gyakran van szükség két pont távolságának a meghatározására. Kis hálózatokban (például ha a 16. ábrán látható hálózatról van szó) ez könnyű feladat. Egy több millió pontból álló hálózatban viszont meglehetősen sok időbe telhet megtalálni a legrövidebb utat. A gyakorlatban erre általában a szélességi keresés (angolul breadth first search) algoritmusát szokták használni.
- *Átmérő:* egy hálózat átmérője (a jele $d_{\max}$) a legrövidebb úthosszak maximuma. Másként fogalmazva: $d_{\max}$ az összes pont–pont távolság legnagyobbika. Ellenőrizhető, hogy a 16. ábrán bemutatott hálózat átmérője $d_{\max} = 3$.
- *Átlagos úthossz, átlagos távolság:* az átlagos úthossz (vagy más néven az átlagos távolság, amelynek jele $\langle d \rangle$) a hálózat pont–pont távolságainak átlaga. Egy N csomópontból áll hálózatban

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{ij} \qquad 7.$$

A kis világ tulajdonság

A *kis világ* jelenség, más néven hatlépésnyi távolság már régóta izgatja a közvéleményt. Arról a kijelentésről van szó, hogy ha bárhol a Földön kiválasztunk valakit, és a Föld bármely részén valaki mást, akkor legfeljebb hat ismerősön át út vezet közöttük. Az egyáltalán nem meglepő, hogy az egy városban lakók között csak néhány kézfogásnyi a távolság. A kis világ elv azt állítja, hogy két ember, még ha a földgolyó átellenes két részében él is, biztosan összeköthető néhány ismerősön át.

Elsőként Karinthy Frigyes írta le a kis világ jelenségét 1929-ben megjelent Láncszemek című elbeszélésében.[12] *Kísérleti téren Stanley Milgram amerikai szociálpszichológus nevéhez*

[11] Adott hosszúságú utak száma és a szomszédsági mátrix. Ha arra vagyunk kíváncsiak, hogy hány darab d hosszúságú út vezet i-ből j-be, ezt könnyűszerrel meghatározhatjuk a szomszédsági mátrix segítségével. $d = 1$ esetén ez triviálisan $N_{ij}^{(d=1)} = A_{ij}$. A 2 hosszúságú utak esetén $N_{ij}^{(d=2)} = \Sigma_q A_{iq} A_{qj} = [A^2]_{ij}$, ahol A^2 az a négyzetes mátrix, amit a szomszédsági matrix önmagával vett szorzatából kapunk. Könnyen belátható, hogy általános d hosszúságú utak esetén $N_{ij}^{(d)} = [A^d]_{ij}$.

[12] KARINTHY 1929.

fűződik ebben a témában az első mérés,[13] *aki 1967-ben „célszemélyként" kiválasztott egy bostoni tőzsdeügynököt és egy sharoni teológushallgatót (mindkét város Massachusetts államban van). Ezután két másik városba, Wichitába és Omahába küldött levelet két véletlenszerűen kiválasztott személynek a kutatás rövid céljával, a célszemély fotójával, nevével, címével és egyéb adatokkal. Arra kérte a címzetteket, hogy küldjék tovább a levelet egy olyan barátjuknak, rokonuknak vagy ismerősüknek, aki valószínűleg ismeri a célszemélyt.*

Néhány nap múlva – mindössze két lépéssel – megérkezett az első levél. A 296 levélből végül 64 érkezett vissza, néhány majdnem egy tucat lépésen át. Ezeknek a befejezett láncoknak az ismeretében Milgram meghatározhatta, hogy hány ember kell egy levél célba juttatásához. Arra jutott, hogy a közvetítők átlagos száma 5,5, ami eredményként később „a hat lépés távolság"-ként vált széles körben ismertté.

Milgramnak nem volt térképe a teljes ismeretségi hálózatról, ezért kísérlete nem tárhatta fel a résztvevők tényleges távolságát. Ma a Facebooknak van a valaha összegyűjtött legátfogóbb társadalmi térképe. A Facebook 2011. májusi ismeretségi gráfját felhasználva (721 millió aktív felhasználóval és 68 milliárd szimmetrikus kapcsolattal) a kutatók megállapították, hogy a felhasználók közötti átlagos távolság 4,74.[14]

A hálózattudomány nyelvén a kis világ jelensége annyit tesz, hogy a hálózatban két véletlenül kiválasztott csomópont között is rövid a távolság. Ez a megállapítás két kérdést vet fel. Mit jelent itt a rövid (vagy kicsi), vagyis mihez képest az? Mivel magyarázzuk továbbá a kis távolságok létezését?

Egyszerű számítással választ adhatunk a fenti két kérdésre. Vegyünk egy nagy (de véges) méretű, $\langle k \rangle > 1$ átlagos fokszámú véletlen gráfot. Tegyük fel, hogy a hálózat ritka, azaz $\langle k \rangle \ll N$. Vizsgáljuk meg, hogy miként alakul egy véletlenszerűen választott pont első-, másod-, harmad- és általánosan d-edrendű szomszédainak száma:

- az első szomszédok száma megegyezik a fokszámmal, $N(d = 3) \simeq \langle k \rangle$;
- a másodszomszédok számát becsülhetjük úgy, mint $N(d = 3) \simeq \langle k \rangle(\langle k \rangle - 1) \simeq \langle k \rangle^2$, hiszen minden első szomszéd átlagosan $\langle k \rangle - 1$ további, az első körben még nem érintett ponthoz vezető kapcsolattal rendelkezik. Itt természetesen az egyszerűség kedvéért elhanyagoltuk annak lehetőségét, hogy két első szomszéd között is lehet él;
- a harmadszomszédok számát hasonló érveléssel úgy közelíthetjük mint $N(d = 3) = \langle k \rangle^3$;
- és általánosan, ezen gondolatmenet szerint általánosan a d-edrendű szomszédok becsült átlagos száma $N_d = \langle k \rangle^d$.

Vegyük észre, hogy az iménti becslésekben az egyre távolabbi szomszédok száma *exponenciálisan* nő a lépésszámmal, ami egy nagyon gyors növekedési ütemet jelent. Természetesen az adott d távolságra lévő pontok száma nem nőhet a teljes hálózat mérete fölé, ezért kell legyen egy maximális d_m távolság, amelynél $N(d_m) \simeq \langle k \rangle^{d_m} \simeq N$. Ez a maximális távolság egy egyszerű becslésként szolgálhat az átlagos távolságra $\langle d \rangle \sim d_m$, amelyet ezek alapján a

[13] MILGRAM 1967.

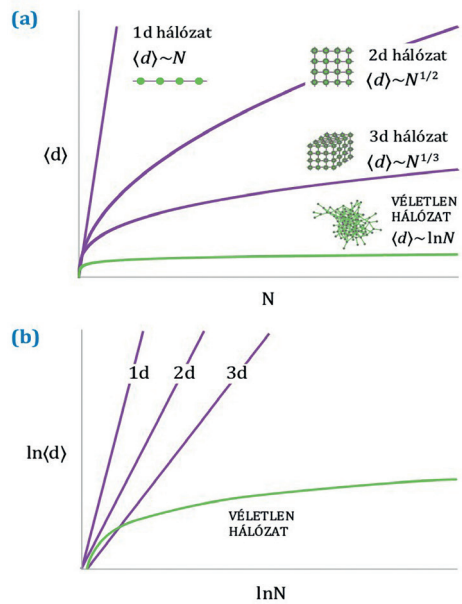
[14] BACKSTROM–BOLDI–ROSA–UGANDER–VIGNA 2012.

$$\langle d \rangle \simeq \frac{\ln N}{\ln \langle k \rangle} \qquad 8.$$

alakban adhatunk meg.

A 8. egyenlet a kis világ jelenségének matematikai alakja. Az, hogy az átlagos távolság a pontok számának logaritmusával arányos – $\langle d \rangle \sim \ln N$ –, nagyon jól teljesül az Erdős–Rényi-modellben (hiszen a fenti, közelítő levezetés is egy ilyen véletlen gráfra vonatkozott), ezért bátran mondhatjuk, hogy a klasszikus véletlen gráf rendelkezik a kis világ tulajdonsággal.

A 8. egyenletben is látható logaritmikus függésről azt érdemes megjegyezni, hogy általában $\ln N \ll N$, tehát a $\langle d \rangle$ értékének $\ln N$-től való függése azzal jár, hogy egy kis világ tulajdonságú hálózatban a távolságok nagyságrendekkel kisebbek a hálózat méreténél. Következésképpen a „kis világ jelensége" kifejezésben „kis"-en azt értjük, hogy az átlagos úthossz logaritmikusan függ a hálózat nagyságától. A „kicsi" tehát azt jelenti, hogy $\langle d \rangle$ az $\ln N$-nel arányos, és nem N-nel vagy N-nek valamilyen hatványával (17. ábra).



17. ábra

A $\langle d \rangle$ a rendszerméret függvényében

Megjegyzés: A távolságokkal kapcsolatos benyomásaink nem kis részben a szabályos rácsokkal szerzett tapasztalatokon alapulnak, ugyanakkor azokban nem bukkan fel a kis világ jelensége. Például egy 1 dimenziós rácsban (N hosszúságú lánc) az átlagos távolság N-nel lineárisan nő: $\langle d \rangle \sim N$. Egy négyzetrácsban $\langle d \rangle \sim N^{1/2}$, egy köbös rácsban $\langle d \rangle \sim N^{1/3}$ és általánosságban, egy D dimenziós rácsban $\langle d \rangle \sim N^{\frac{1}{D}}$, és ezek a polinomiális függvénykapcsolatok a 8. egyenlettel összefüggésben látottnál gyorsabb növekedést jeleznek, rácsban az úthosszak tehát jóval nagyobbak, mint véletlen hálózatban. Ezt a jelenséget szemlélteti lineáris skálán az a) panel, és logaritmikus skálán a b) panel.

Forrás: BARABÁSI 2017

A kérdés ezek után az, hogy a valódi komplex rendszereket reprezentáló hálózatokban az átlagos úthossz vajon miként viselkedik? A legtöbb esetben ugyan konkrét méretfüggést nem tudunk vizsgálni, de azt feltétlen meg lehet állapítani, hogy vajon a pontok száma és az átlagos úthossz között nagyságrendi eltérés tapasztalható, avagy sem.

3. táblázat

A mintaként használt tíz hálózat átlagos úthossza

Hálózat neve	N	L	$\langle d \rangle$	$\langle d_{ER} \rangle$
Internet	192 244	609 066	6,98	6,58
WWW	325 729	1 497 134	11,27	8,31
Áramellátás	4 941	6 594	18,99	8,66
Mobilhívások	36 595	91 826	11,72	11,42
E-mail	57 194	103 731	5,88	18,4
Tudományos együttműködések	23 133	93 439	5,35	4,81
Színészek hálózata	702 388	29 397 908	3,91	3,04
Hivatkozási hálózat	449 673	4 689 479	11,21	5,55
E. coli anyagcseréje	1 039	5 802	2,98	4,04
Élesztő fehérjéinek kölcsönhatásai	2 018	2 930	5,61	7,14

Megjegyzés: A lemért átlagos úthossz a 4. oszlopban szerepel. Összehasonlításként az 5. oszlop az adott hálózattal egyező méretű és élsűrűségű klasszikus véletlen gráf átlagos úthosszát mutatja.

Forrás: BARABÁSI 2017

A 3. táblázatban a mintaként használt tíz hálózatban mérhető átlagos úthosszakat soroltuk fel. Látható, hogy majdnem minden esetben legalább három nagyságrendű eltérés tapasztalható $\langle d \rangle$ és N között, és $\langle d \rangle$ minden esetben húsz alatt marad annak ellenére is, hogy olyan hálózat is szerepel a listában, amely több mint félmillió pontból áll. Összehasonlításként feltüntettük a felsorolt valós rendszerekkel egyező méretű és élsűrűségű Erdős–Rényi-gráf átlagos úthosszát is, amely nagyon jó egyezést mutat a valódi hálózat értékével a legtöbb esetben. Ezek alapján bátran kijelenthetjük, hogy *a valódi hálózatok is rendelkeznek a kis világ tulajdonsággal.*

Ez a megállapítás nagyon szemléletes például az emberi ismeretségek hálózata esetén. Ha az ismeretségek hálózata egy kétdimenziós négyzetrácsot alkotna, és mindenki csak a közvetlen szomszédját ismerné, akkor két ember közötti távolság átlagosan nagyjából $(7\times10^9)^{1/2} = 83{,}666$ lenne. Még ha ezt ki is javítjuk arra, hogy egy embernek nem négy, hanem nagyjából 1000 ismerőse van, az átlagos távolságra még akkor is nagyságrendekkel nagyobb érték jönne ki annál, amennyit a 8. egyenlet ad.

Klaszterezettségi együttható

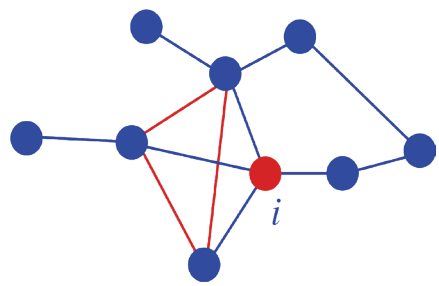
Tegyük fel magunknak a következő kérdést: ha az emberi ismeretségek hálózatából kiválasztjuk két közvetlen ismerősünket, akkor ők vajon egymást is ismerik? Természetesen ez attól függ, hogy konkrétan melyik két ismerősünket választottuk, de az valószínűleg mindenkire igaz, hogy a válasz sok esetben „igen” lesz, és az „igen” válaszok hányada

lényegesen magasabb, mint ha véletlenszerűen választanánk két embert az egész hálózatból. Ez a gondolatkísérlet arra próbál rávilágítani, hogy az ismeretségi hálózatban sok háromszög alakul ki, amelyekben egy adott ember két ismerőse egymással is kapcsolatban van.

Általánosságban egy csomópont környezetében a háromszögek előfordulási gyakoriságának jellemzésére vezették be a *klaszterezettségi együtthatót,* amelyet szokás csoporterősségi együtthatónak is nevezni. Ennek értéke mindig 0 és 1 közé esik, és megegyezik az adott csúcs közvetlen szomszédai között fellelhető élek és a szomszédok közt behúzható élek maximális számának hányadosával. A továbbiakban az i pont klaszterezettségi együtthatóját C_i-vel fogjuk jelölni, és az iménti definíció alapján egy irányítatlan hálózatban a C_i a következő formulával adható meg:

$$C_i = \frac{2e_i}{k_i(k_i - 1)}, \qquad 9.$$

ahol ei az i szomszédai közt fellelhető élek száma, és kihasználtuk, hogy az i-hez tartozó k_i darab közvetlen szomszéd között maximálisan $k_i(k_i - 1)/2$ kapcsolat volna lehetséges. (A C_i kiszámítását egy egyszerű példán keresztül a 18. ábrán szemléltetjük.)



18. ábra

A klaszterezettségi együttható meghatározása

Megjegyzés: Az ábrán pirossal jelzett *i* csúcs fokszáma $k_i = 4$, és a szomszédai között összesen 3 él található (ezeket is piros színnel emeltük ki). Ennek alapján az *i* klaszterezettségi együtthatója $C_i = \frac{2 \cdot 3}{4 \cdot 3} = \frac{1}{2}$.

Forrás: a szerző szerkesztése

Irányított esetben természetesen a lehetséges kapcsolatok száma megkétszereződik (a páronkénti két lehetséges irány miatt), ennélfogva irányított hálózatokban:

$$C_i = \frac{e_i}{k_i(k_i - 1)}. \qquad 10.$$

Joggal merülhet fel a kérdés, hogy mi a helyzet azokkal a csúcsokkal, amelyeknek csak egy közvetlen szomszédjuk van ($k_i = 1$), vagy teljesen izoláltak ($k_i = 0$). A 9. és 10. egyenletekben ilyenkor 0-val kéne osztani, ami matematikailag problémás. Az ilyen esetekre való tekintettel a klaszterezettségi együttható értéke a $k_i \leq 1$ fokszámú csúcsokra definíció-szerűen nulla, azaz $C_i = 0$.

Vegyük észre, hogy a klaszterezettségi együttható definíciója nagyon hasonlít az élsűrűség 4. egyenletben ismertetett definíciójához, hiszen itt is bizonyos csúcsok között létező élek számának és ugyanazon csúcsok között lehetséges élek számának a hányadosával van dolgunk. Sőt, ha a teljes hálózat helyett pusztán az közvetlen szomszédai között kifeszülő hálózatot vizsgáljuk (magát az i csúcsot kihagyva), akkor ennek az új (vélhetően sokkal kisebb) hálózatnak az élsűrűsége pontosan megegyezik a C_i-vel. Ennek révén a klaszterezettségi együtthatóra gondolhatunk úgy is, mint egyfajta *lokális élsűrűségre,* amely megadja, hogy az i pont véletlenszerűen választott két szomszédja között mekkora valószínűséggel találunk közvetlen kapcsolatot.

Átlagos klaszterezettségi együttható

Az egyes pontokon mérhető klaszterezettségi együtthatókat átlagolva kapjuk meg a hálózat átlagos klaszterezettségi együtthatóját:

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^{N} C_i \qquad 11.$$

Az iménti felismerés alapján erre tekinthetünk úgy, mint egyfajta *átlagos lokális élsűrűségre,* amely megadja, hogy egy véletlenszerűen választott pont közvetlen szomszédai egymással mekkora valószínűséggel vannak összekötve. De természetesen gondolhatunk $\langle C \rangle$-re úgy is, mint a lokális „tranzitivitásra" jellemző mennyiségre.

Vizsgáljuk meg ezek után az Erdős–Rényi-féle klasszikus véletlen gráf átlagos klaszterezettségi együtthatóját. Mivel ebben a modellben bármely két pont között minden mástól függetlenül p valószínűséggel találunk kapcsolatot, az átlagos lokális élsűrűség és a globális, teljes hálózatra vonatkozó élsűrűség megegyezik. Másként megfogalmazva, ha egy Erdős–Rényi-gráfban választunk egy pontot véletlenszerűen, majd megvizsgáljuk ezen pont szomszédai között a kapcsolatokat, akkor átlagosan azt fogjuk tapasztalni, hogy épp annyira valószínű ezen szomszédok között az élek felbukkanása, mint bármely két pont között a hálózatban. Ezek alapján az Erdős–Rényi-gráf átlagos klaszterezettségi együtthatója megegyezik az élbekötési valószínűséggel, $\langle C \rangle = p$.

A valós hálózatok magasan klaszterezettek

Térjünk rá most a valódi komplex rendszereket leíró hálózatok klaszterezettségi együtthatójára. A 4. táblázatban a korábban már bemutatott tíz hálózat (1. táblázat) átlagos klaszterezettségi együtthatóját soroljuk fel a 4. oszlopban. Első ránézésre változatos értékek fordulnak elő, amelyek így önmagukban nem mutatnak egységes tendenciát.

4. táblázat

A mintaként használt tíz hálózat átlagos klaszterezettsége

Hálózat neve	N	L	$\langle C \rangle$	$\langle C_{\mathrm{ER}} \rangle$
Internet	192 244	609 066	0,180	$3{,}30 \times 10^{-5}$
WWW	325 729	1 497 134	0,108	$1{,}41 \times 10^{-5}$
Áramellátás	4 941	6 594	0,08	$5{,}40 \times 10^{-4}$
Mobilhívások	36 595	91 826	0,141	$6{,}86 \times 10^{-5}$
E-mail	57 194	103 731	0,036	$3{,}17 \times 10^{-5}$
Tudományos együttműködések	23 133	93 439	0,726	$3{,}49 \times 10^{-4}$
Színészek hálózata	702 388	29 397 908	0,790	$1{,}19 \times 10^{-4}$
Hivatkozási hálózat	449 673	4 689 479	0,240	$2{,}32 \times 10^{-5}$
E. coli anyagcseréje	1 039	5 802	0,377	0,00538
Élesztő fehérjéinek kölcsönhatásai	2018	2930	0,046	0,00144

Megjegyzés: A lemért átlagos klaszterezettségi együtthatót az 4. oszlopban tüntettük fel. Összehasonlításként az 5. oszlop az adott hálózattal egyező méretű és élsűrűségű klasszikus véletlen gráf átlagos klaszterezettségét mutatja.

Forrás: BARABÁSI 2017

Azonban ha egyesével összehasonlítjuk a mért értéket az 5. oszlopban látható értékkel, amely az azonos méretű és élsűrűségű Erdős–Rényi-gráf átlagos klaszterezettségét tünteti fel, egyből világossá válik, hogy *a valós hálózatok átlagos klaszterezettsége sokkal magasabb, mint a nekik megfelelő klasszikus véletlen gráfé.* Ezt másként úgy fogalmazhatjuk meg, hogy bár a valós hálózatok globálisan ritkák (lásd *A valós hálózatok ritkák* című alfejezetet), lokálisan mégis sokkal sűrűbbnek tűnnek, mint azt az ember egy véletlengráf-szerű megközelítésben várná mondjuk a mért $\langle k \rangle$ alapján.

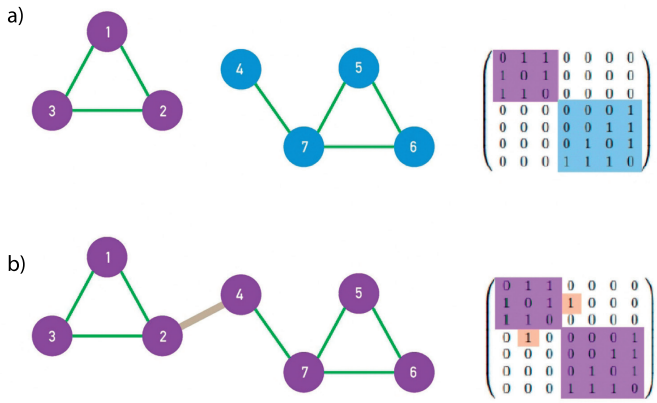
Komponensek és összekötöttség

A telefon igencsak korlátozott kommunikációs eszköz lenne, ha nem hívhatnánk fel vele bármilyen létező telefonszámot. Az e-maillel sem mennénk sokra, ha csak meghatározott címekre küldhetnénk e-mailt. A hálózatkutatás szempontjából mindez azt jelenti, hogy a telefonálást és az internetet megtestesítő hálózatokban kell hogy legyen út bármely két pont között. Voltaképp ez a legtöbb hálózat legfőbb haszna: *összekapcsoltsággal* szolgál. Ebben a részben az *összekötöttség* gráfelméleti megfogalmazásával foglalkozunk.

Egy irányítatlan hálózatban az *i*-edik és a *j*-edik pont egymással *összekötött* akkor, ha a hálózatban létezik közöttük út, és egymással *összekötetlen,* ha a kettőt nem köti össze út; ez utóbbi esetben $d_{ij} = \infty$. Egy hálózat akkor *összefüggő,* ha bármelyik két pontja összekötött, azaz van köztük a hálózat éleiből álló út. A hálózat nem összefüggő, ha van legalább két olyan pontja, amelyek között nincsen út. Ilyenkor azt mondjuk, hogy a hálózat több *komponensből* áll.

A hálózatban egy *összefüggő komponens* (vagy röviden csak komponens) egy olyan maximális részgráf, amelyen belül bármely két pont között van út. A „maximális" itt azt

jelenti, hogy egyetlen további pontot sem tehetünk hozzá a hálózatból úgy, hogy megmaradjon ez a tulajdonsága.



19. ábra

Nem összefüggő és összefüggő hálózatok

Megjegyzés: a) Egy két különálló komponens alkotta kis hálózat. Az (1, 2, 3) komponensen belül és a (4, 5, 6, 7) komponensen belül bármely pontpár között van út. Két különböző komponenshez tartozó pontok között azonban nincs. Az ábra jobb oldalán látható a hálózat szomszédsági mátrixa. Ha a hálózat több komponensből áll, akkor a szomszédsági mátrix blokkdiagonális alakra hozható: a főátlóra (a mátrix bal felső sarkától a jobb alsó sarkáig tartóra) eső négyzetes blokkok kivételével minden mátrixelem nulla lesz. b) Egy hídnak nevezett kapcsolat hozzáadásával (az ábrán ezt a kapcsolatot szürke szín jelöli) itt egy nem összefüggő hálózatból egyetlen összefüggő komponenst kapunk. Most már minden pontpár között van út, és a szomszédsági mátrix nem alakítható át a teljes mátrixnál kisebb blokkokból álló blokkdiagonális alakúra.

Forrás: BARABÁSI 2017

Szemléltetésül a 19. ábra a) képén egy két komponensből álló hálózatot mutatunk a neki megfelelő szomszédsági mátrixszal együtt.

Ha egy hálózat két komponensből áll, akkor egyetlen, megfelelő helyen létesített kapcsolat összekapcsolt hálózattá teheti [19. ábra b) képe]. Az ilyen kapcsolatot *hídnak* nevezzük. Általánosságban elmondható, hogy a híd a hálózatnak olyan éle, amelynek az elvágása után a hálózat megszűnik összefüggőnek lenni.

Komponensek irányított hálózatokban

A *Távolság* című alfejezetben már említettük, hogy egy irányított hálózatban az a tény, hogy az *i*-edik pontból van út a *j*-edik pontba, nem jelenti automatikusan azt, hogy visszafelé is elérhető *i* a *j*-ből indulva. Emiatt irányított hálózatokban kétfajta komponenst különböztethetünk meg egymástól:

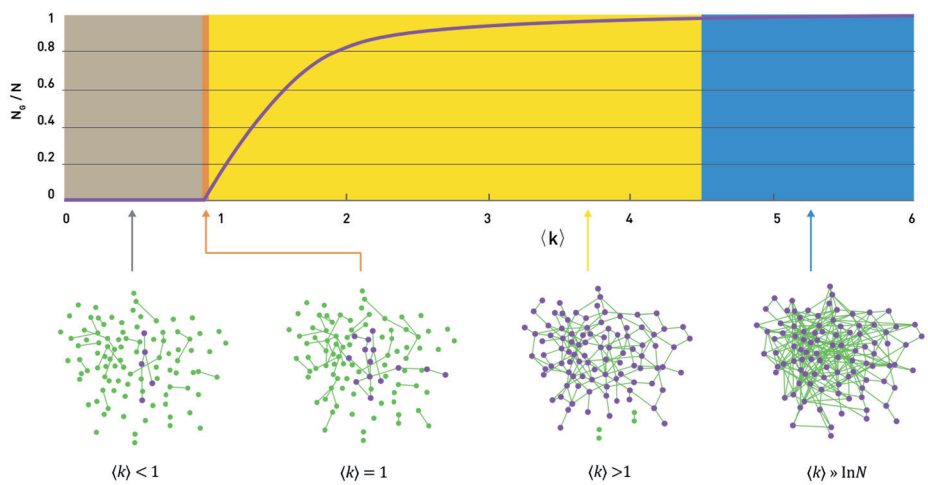
- *Erősen összefüggő komponens* (vagy erősen összekötött komponens): ez olyan maximális részgráfnak felel meg, amelyben bármely két pont között mindkét irányban van út.
- *Gyengén összefüggő komponens* (vagy gyengén összekötött komponens): ez pedig egy olyan maximális részgráf, amelyben bármely két pont között legalább az egyik irányban van út. A gyengén összefüggő komponensek megfelelnek az élek irányának elhagyásával kapott irányítatlan hálózat „sima” komponenseinek.

Óriás összefüggő komponens

Egy hálózat funkcionalitásáról sokat elárulhat a benne található legnagyobb összefüggő komponens mérete. Mint korábban említettük, egy hálózat legalapvetőbb feladata az, hogy közvetlen vagy közvetett elérést biztosítson a csomópontoknak egymás között. Amennyiben a legnagyobb komponens pontjainak N_G száma közelít N-hez, a hálózat „jól teljesít" ezen a téren, hiszen a csúcsok döntő többségéből el lehet jutni majdnem az összes többi pontba.

Mivel mind N, mind N_G széles határok között mozoghat, általában a legnagyobb komponens relatív méretét szokás vizsgálni, amely az $S_G = N_G/N$ alakban adható meg, és garantáltan 0 és 1 közötti értéket vesz fel. Elsőként röviden vizsgáljuk meg, hogy miként viselkedik S_G az Erdős–Rényi-modellben. Ha fixáljuk a pontok N számát, akkor ebben a modellben már csak a p élbekötési valószínűséget tudjuk változtatni. A $p = 0$ határesetben egyáltalán nem húzunk be élt a pontok között, ennélfogva a hálózat N darab 1 méretű komponensből fog állni, azaz a legnagyobb komponens relatív mérete $S_G = 1/N$. Ellenkező esetben, ha a maximálisan lehetséges $p = 1$ értéket választjuk, akkor az összes lehetséges élt létrehozzuk a hálózatban, amely ennélfogva 1 darab $N_G = N$ méretű komponensből fog állni, azaz ilyenkor $S_G = 1$.

Ha fokozatosan változtatjuk a p értékét a két extrém határeset között, akkor egy érdekes átalakulást figyelhetünk meg S_G-ben, amelyet a 20. ábra mutat be.



20. ábra

A legnagyobb komponens relatív mérete a klasszikus véletlen gráfban

Megjegyzés: Az ábra felső részében az $S_G = N_G/N$ értékét ábrázoltuk a $\langle k \rangle = (N-1)p$ átlagos fokszám függvényében. Alatta szemléltetésül egy-egy kis méretű véletlen gráfot láthatunk a különböző átlagos fokszám tartományokból. Ezekben a legnagyobb komponenshez tartozó pontokat lila színnel jelöltük.

Forrás: BARABÁSI 2017

A grafikon függőleges tengelye természetesen az $S_G = N_G/N$-nek felel meg, a vízszintes tengelyen azonban a p helyett a $\langle k \rangle = (N-1)p$ átlagos fokszám szerepel. Ennek oka az,

hogy a különböző N mérethez tartozó Erdős–Rényi-gráfokra vonatkozó görbék így lényegében egy univerzális görbére „húzódnak” rá. Az ábrán a végtelen nagy, $N \to \infty$ véletlen gráf görbéje szerepel; a véges, de már kellően nagy Erdős–Rényi-gráfok görbéje ehhez már nagyon közel esik.

A 20. ábrán látható grafikon alapján leszűrhető, hogy amennyiben $\langle k \rangle < 1$, a legnagyobb komponens relatív mérete 0-hoz tart; ez azt jelenti, hogy ebben a tartományban a véletlen gráf túl ritka ahhoz, hogy egy számottevő összefüggő komponens ki tudjon benne alakulni. Azonban ahogy $\langle k \rangle$ eléri a $\langle k \rangle = 1$ értéket, egy drasztikus átalakulást tapasztalhatunk. A legnagyobb összefüggő komponens mérete elkezd meredeken emelkedni a $\langle k \rangle$ függvényében. Mivel az ábrán a végtelen nagy véletlen gráfra vonatkozó görbét láthatjuk, ebből az következik, hogy a $\langle k \rangle > 1$ tartományon a legnagyobb komponenshez is már végtelen sok pont tartozik. Az ilyen, végtelen nagy méretű komponenst hívjuk óriás összefüggő komponensnek vagy röviden óriás komponensnek. Ha $\langle k \rangle$ értékét még tovább növeljük, eljutunk egy olyan tartományba, ahol az óriás komponens már lényegében az egész hálózatra kiterjed, és S_G kezd az 1-hez közelíteni.

Ha visszatérünk ahhoz a gondolathoz, hogy egy hálózat alapfeladata a pontok összekötöttségének biztosítása, akkor az óriás komponens léte vagy hiánya a hálózatban egy ezzel kapcsolatos nagyon fontos indikátor (hiszen, ha van óriás komponens, akkor a pontoknak egy nem elhanyagolható hányada el tudja érni egymást). Ennek fényében a 20. ábra grafikonján látható, $\langle k \rangle = 1$-nél történő átalakulás rendkívül fontos, hiszen ezen a ponton (amit szokás kritikus pontnak is nevezni) jelenik meg a hálózatban (vagy tűnik el a hálózatból a másik irányban haladva) az óriás komponens.

Térjünk rá ezek után röviden arra, hogy vajon a komplex rendszereket reprezentáló valós hálózatok tartalmaznak-e óriás komponenst? Noha ezek a hálózatok mindig véges méretűek, és ezért szigorú értelemben nem találhatunk bennük végtelen sok pontot tartalmazó komponenst, a gyakorlati tapasztalat azt mutatja, hogy minden rendszer esetén rendelkeznek egy olyan komponenssel, amelynek mérete összevethető a csúcsok teljes számával. Ennélfogva a valós hálózatokra tekinthetünk úgy is, mint óriás összefüggő komponenssel rendelkező hálózatokra. Mindazonáltal gyakori, hogy egy valós hálózat óriás komponense nem terjed ki minden pontra, és bennük az óriás komponens mellett még további, sokkal kisebb (az óriás komponenstől izolált) komponenseket is találunk.

Centralitás

Egy komplex hálózat elemzése során gyakran merül fel az a kérdés, hogy melyek a legfontosabb pontok a hálózatban, és miként lehet a pontokat rangsorolni? Természetesen erre nincs általános eljárás, mindig az adott probléma határozza meg, hogy milyen szempontból számíthat fontosnak egy-egy csomópont. Azonban megadható néhány mennyiség, ami az esetek döntő többségében hasznos támpontot nyújthat a fontos, központi csomópontok megkereséséhez. Ezeket a mennyiségeket hívjuk *centralitásoknak.* Az egyik legegyszerűbb centralitás maga a fokszám, angolul *degree centrality.* A továbbiakban néhány, széles körben használt centralitást vázolunk röviden.

Közelség

A *közelség* (angolul *closeness,* illetve *closeness centrality*) a csomópontok többi ponttól vett távolságát veszi figyelembe, és ez alapján rangsorol. Ennél a centralitásnál az az alapgondolat, hogy a hálózat „középső", centrális régiójában elhelyezkedő csúcsok átlagosan közelebb vannak az összes többi csúcshoz, mint a hálózat perifériáján található pontok.

Ez alapján először bevezethetjük egy *i* csúcs átlagos távolságát a többi csúcstól a

$$\langle d_i \rangle = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} d_{ij} \qquad 12.$$

alakban. Ahhoz, hogy ebből egy olyan centralitást kapjunk, amely azon csúcsoknál vesz fel nagy értéket, amelyek az átlagosnál közelebb vannak a többi csúcshoz, a kifejezést még invertálni kell, ami alapján az *i* közelségét a

$$C_c(i) = \frac{1}{\langle d_i \rangle} \qquad 13.$$

alakban szokás definiálni.

Vegyük észre, hogy ha a hálózat több komponensből áll, akkor gondban leszünk a 12. és 13. egyenletek kiértékelésével, hiszen bármely ponttól a vele nem azonos komponensben lévő másik pontokhoz mért távolság $d_{ij} = \infty$, s ennélfogva minden pontra $\langle d_i \rangle = \infty$ és $C_c(i) = 0$ eredményt kapunk. Emiatt a több komponensből álló hálózatok esetén módosítani kell a 12. és 13. képleteket úgy, hogy az átlagos távolságot csak az azonos komponenshez tartozó csúcsok figyelembevételével számoljuk ki. Ezzel elkerüljük a végtelen távolságok felbukkanását, de ugyanakkor aránytalanul nagy előnyhöz juttatjuk a kis komponensekhez tartozó pontokat, hiszen például egy összesen két pontból (és az őket összekötő élből) álló komponens esetén $\langle d_i \rangle = d_{ij} = 1$, és emiatt $C_c(i) = 1$ jön ki, amivel egyszerűen nem versenyezhet például egy óriás komponenshez tartozó csúcs, bármennyire is centrálisan helyezkedik el.

Ezt a komponensméretből adódó torzítást úgy szokás kiküszöbölni, hogy a közelség definíciójában figyelembe vesszük az adott csúcs komponensének relatív méretét is. Jelöljük az i-edik csúcs komponensét $K(i)$-vel, és az ehhez tartozó pontok számát $N_K(i)$-vel. A fentiek alapján az i pont $C_c(i)$ közelségét a

$$\langle d_i \rangle = \frac{1}{N_K(i)-1} \sum_{\substack{j \in K(i) \\ j \neq i}} d_{ij} \qquad 14.$$

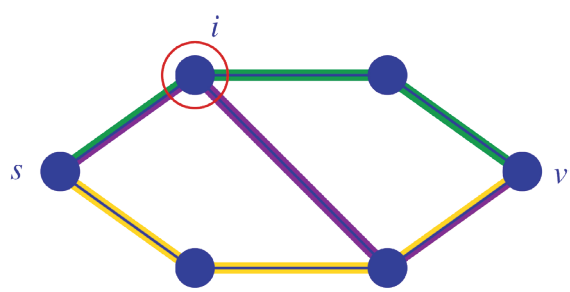
$$C_c(i) = \frac{N_K(i)}{N} \frac{1}{\langle d_i \rangle} \qquad 15.$$

képletek alapján határozhatjuk meg. Érdemes megjegyezni, hogy ha a 14. és 15. egyenleteket egy olyan hálózatra alkalmazzuk, amelyben csak 1 komponens van, [azaz $N_K(i) = N$, és minden pont a $K(i)$-hez tartozik], akkor pontosan visszakapjuk a közelség eredeti definícióját a 12. és 13. egyenletekben megadott alakban.

Köztesség

A komponensek tárgyalásánál már szóba kerültek a hidak, amelyek nagyon fontos összekötő szerepet játszanak, és elvételükkel egy korábban összefüggő komponens két (a továbbiakban már nem összekötött) részre esik szét. Egy ilyen híd kiemelt szerepe abban is tükröződik, hogy nagyon sok csúcspár között a legrövidebb útnak át kell haladnia rajta, hiszen ha bármely pontból indulva a híd egyik végéhez kapcsolódó részgráfon el akarunk jutni a híd másik végéhez kapcsolódó részbe, akkor az útnak át kell haladnia a hídon.

Ha esetleg egy nagy hálózatban szeretnénk automatizáltan olyan éleket vagy pontokat keresni, amelyek esetleg híd szerepet tölthetnek be, a fentiek alapján egy kézenfekvő gondolat azt vizsgálni, hogy hány legrövidebb út halad át összesen az adott élen vagy ponton. Ez vezet el minket a *köztesség* (angolul *betweenness* vagy *betweenness centrality*) definíciójához, amely az adott pontra (vagy élre) nézve a rajta áthaladó legrövidebb utak súlyozott összege, ahol a súlyozás azt veszi figyelembe, hogy egy adott csúcspár között több legrövidebb út is létezhet, és ezek nem feltétlenül haladnak át mind az éppen vizsgált ponton. Szemléltetésül tekintsük a 21. ábrát.



21. ábra

A köztesség meghatározása

Megjegyzés: Ha az i pont köztességét vizsgáljuk, akkor az (s,v) pontpár közötti legrövidebb utak járulékánál a következők szerint kell eljárnunk: összesen 3 legrövidebb út található *s* és *v* között, a zöld, lila és sárga színekkel jelzett utak. Ezek közül 2 halad át konkrétan az *i* csomóponton (a zöld és a lila), ezért az *i* köztességében az (s,v) pontpár járuléka ⅔.

Forrás: a szerző szerkesztése

Tegyük fel, hogy a mutatott hálózatban az i pont köztességét szeretnénk meghatározni. Ehhez végig kell mennünk az összes olyan pontpáron, amelynek nem tagja i, és meg kell vizsgálnunk, hogy az adott páros között a legrövidebb út vajon áthalad-e az i ponton, vagy sem. Amennyiben a legrövidebb út érinti az i pontot, úgy az i köztességéhez hozzáadunk 1-et, ha elkerüli, akkor nem adunk hozzá semmit. Igen ám, de ahogy a 21. ábrán az (s, v)

páros esetén láthatjuk, a helyzetet bonyolíthatja, ha az adott pontpár között több legrövidebb út is létezik párhuzamosan. Ilyenkor, ha ezen legrövidebb utak közül valahány áthalad az i ponton, akkor az i köztességét az áthaladó utak hányadával kell növelni az összes legrövidebb utak számához képest.

Ha a fentiek alapján képletbe szeretnénk önteni a köztesség definícióját, akkor először bevezethetjük az s és v pontok közti legrövidebb utak számára a σ_{sv} jelölést, illetve jelölhetjük $\sigma_{sv}(i)$-vel azon legrövidebb utak számát s és v között, amelyek át is haladnak az i csúcson, és ezek segítségével az i köztessége a

$$C_b(i) = \sum_{\substack{s \neq i \\ v \neq i}} \frac{\sigma_{sv}(i)}{\sigma_{sv}} \qquad 16.$$

alakban adható meg.

Összetett hálózatjellemzők

Ebben a fejezetben olyan hálózatjellemzőket ismertetünk, amelyek már nem az egyes csúcsok különböző mérhető tulajdonságaira koncentrálnak, mint az előző fejezetekben tárgyalt mennyiségek jelentős része, hanem az egész hálózat statisztikus jellemzésére törekszenek, különféle szempontok alapján. Ezek lehetővé teszik különböző hálózatok összehasonlítását, illetve esetenként a hálózatok osztályozását. Természetesen ezek a statisztikák erősen támaszkodnak a korábban tárgyalt mutatókra.

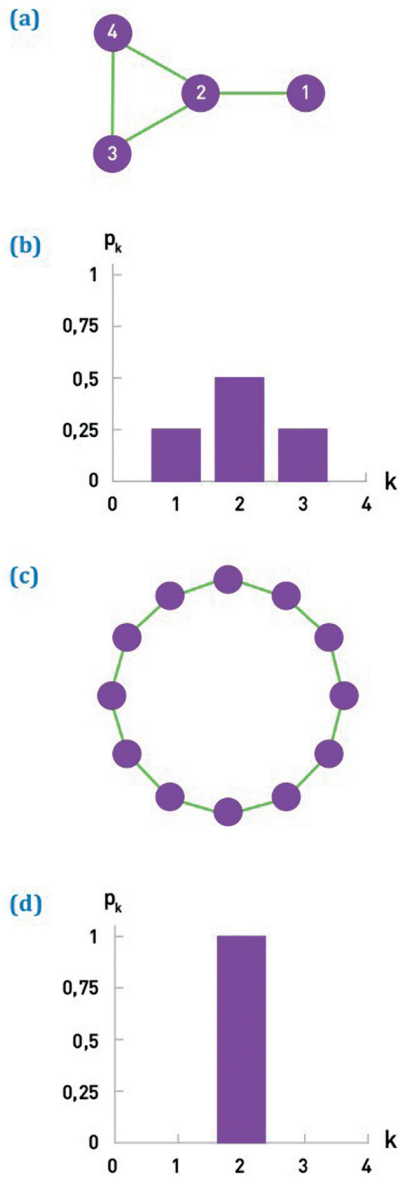
Fokszámeloszlás

A *Fokszám* című alfejezetben megismerkedtünk a fokszám és az átlagos fokszám fogalmával. A statisztikai leírás irányába ezen a téren a következő természetes lépés a hálózatban előforduló fokszámok eloszlásának vizsgálata. Ezt az eloszlást hívjuk *fokszámeloszlásnak*. Definíciószerűen a $p(k)$ fokszámeloszlás értéke adott k fokszámnál annak valószínűsége, hogy egy véletlenszerűen választott pont fokszáma k a hálózatban.

Egy véges méretű, valamilyen létező komplex rendszert reprezentáló valós hálózat esetén a fokszámeloszlás elkészítése technikailag egyáltalán nem bonyolult, hiszen ilyen esetekben $p(k)$ értéke egyszerűen a k fokszámú csúcsok hányada a hálózatban, azaz ha N_k-val jelöljük a k fokszámú csúcsok számát, akkor

$$p(k) = \frac{N_k}{N} \qquad 17.$$

Ez alapján ilyen esetekben a fokszámeloszlást egy normalizált hisztogramként is elképzelhetjük, ahogy azt a 22. ábra szemlélteti.



22. ábra
A fokszámeloszlás

Megjegyzés: A $p(k)$ fokszámeloszlás szemléletes jelentése az, hogy megadja a különböző fokszámok előfordulási gyakoriságait. a) Egy kis méretű szemléltető hálózat. b) Az a) panelhez tartozó hálózat fokszámeloszlása. c) Egy „gyűrű" mint hálózat. d) Az előző panelben mutatott gyűrű fokszámeloszlása.

Forrás: BARABÁSI 2017

A véletlen gráf fokszámeloszlása

Mielőtt továbblépnénk, érdemes röviden megvizsgálni, milyen alakot vesz fel a fokszámeloszlás az Erdős–Rényi-féle (klasszikus) véletlen gráfban.[15] Ebben a modellben bármely két csúcs között p valószínűséggel létezik él, és $1-p$ valószínűséggel nem létezik, ezért ha véletlenszerűen kiválasztunk egy csomópontot, akkor annak valószínűsége, hogy pontosan k kapcsolata lesz, a

$$p(k)=\binom{N-1}{k}p^k(1-p)^{N-1-k} \qquad 18.$$

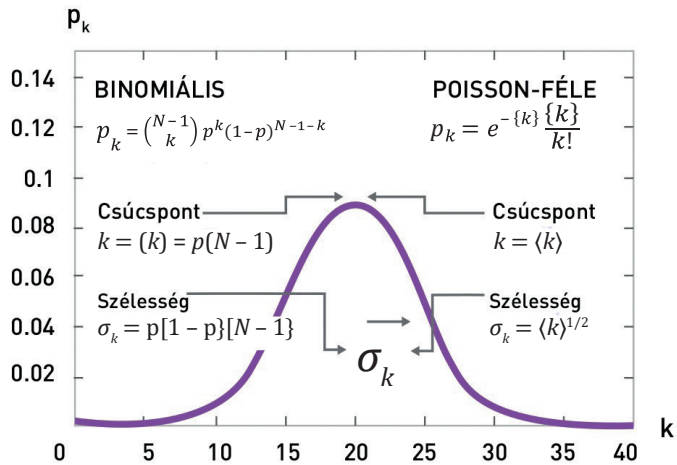
alakban adható meg. A fenti kifejezésben a p^k tényező a csomópont létező kapcsolatai miatt jelenik meg (ahol kihasználtuk az élek függetlenségét, ami miatt a valószínűségek szimplán összeszorzódnak), az $(1-p)^{N-1-k}$ tényező a további $N-1-k$ csomópont felé lehetséges, de most nem létező élekből származik. A kombinatorikai $\binom{N-1}{k}$ tényező azt számolja le, hogy összesen hányféleképpen lehet a k tényleges közvetlen szomszédot kiválasztani az $N-1$ darab összes lehetséges szomszédból.

A 18. kifejezés a valószínűségszámításban gyakran felbukkanó binomiális eloszlás alakját veszi fel, amellyel kapcsolatban a gyakorlatban két fontos közelítéssel szokás élni. Egyrészt nagy számú pontok esetén az $N-1$ és az N között a relatív eltérés elenyészővé válik, ezért gyakori, hogy az egyszerűség kedvéért szimplán N-et írunk $N-1$ helyébe a 18. egyenletben. Másrészt a valószínűségelméletben jól ismert, hogy kellően nagy N esetén a binomiális eloszlás jól közelíthető *Poisson-eloszlással:*

binomiális Poisson

$$p(k)=\binom{N}{k}p^k(1-p)^{N-k} \xrightarrow[\substack{p\to 0\\ Np\to\langle k\rangle}]{N\to\infty} p(k)=\frac{\langle k\rangle^k}{k!}e^{-\langle k\rangle} \qquad 19.$$

[15] Erdős–Rényi 1959.



23. ábra

A binomiális és a Poisson-eloszlás

Megjegyzés: Egy véletlen hálózat fokszámeloszlásának pontos alakja a binomiális eloszlás (baloldalt). Ha $N \gg \langle k \rangle$, akkor a binomiális eloszlás jól közelíthető Poisson-eloszlással (lásd jobb oldalt). Mivel mindkét eloszlás ugyanazt a mennyiséget írja le, ezért ugyanazok a tulajdonságaik, de ezek a tulajdonságok különböző paraméterekkel fejezhetők ki: a binomiális eloszlás N és p függvénye, a Poisson-eloszlásnak viszont csak egyetlen paramétere van: a $\langle k \rangle$. Ez az egyszerűség teszi vonzóvá a Poisson-eloszlás alkalmazását.

Forrás: BARABÁSI 2017

A Poisson-eloszlás gyakorlati előnye a binomiális eloszlással szemben, hogy csak egy paramétere van, amely viszont éppen megegyezik a hálózat $\langle k \rangle = Np$ átlagos fokszámával (23. ábra).

Érdemes még listaszerűen felsorolni a klasszikus véletlen gráf fokszámeloszlásának legfontosabb jellemzőit.

- A fokszámeloszlás *várható értéke* minden esetben megegyezik az átlagos fokszámmal,

$$\langle k \rangle = \sum_k k p(k) \tag{20.}$$

az Erdős–Rényi-gráf esetén (amint a *Fokszám* című alfejezet elején részletesen is tárgyaltuk) ez

$$\langle k \rangle_{\mathrm{ER}} = (N-1)p \simeq Np \tag{21.}$$

alakban adható meg.

- A fokszámeloszlás *szórásnégyzete* vagy más néven *varianciája* definíciószerűen

$$\sigma^2 = \left\langle (k - \langle k \rangle)^2 \right\rangle = \left\langle k^2 \right\rangle - \langle k \rangle^2 = \sum_k k^2 p(k) - \sum_k k p(k) \tag{22.}$$

ami Erdős–Rényi-gráf esetén a

$$\sigma_{\mathrm{ER}}^2 = Np(1-p) \qquad 23.$$

alakban adható meg, ha a binomiális eloszlásból indulunk ki.[16] Amennyiben $N \gg \langle k \rangle = Np$ (azaz a hálózat ritka), úgy szükségszerűen $p \ll 1$ és emiatt $1 - p \simeq 1$, aminek révén

$$\sigma_{\mathrm{ER}}^2 \simeq Np \simeq \langle k \rangle \qquad 24.$$

azaz a szórásnégyzet lényegében megegyezik az átlagos fokszámmal. A szórás (amely szimplán a szórásnégyzet gyökével egyenlő) az eloszlás szélességére, a várható érték körüli fluktuációk nagyságára jellemző.

Skálafüggetlen hálózatok

Térjünk most rá a társadalom és a természet komplex rendszereit reprezentáló hálózatok fokszámeloszlására. Az ezzel kapcsolatos első fontos megjegyzés az, hogy *nem követik a Poisson-eloszlást.* Ha például az emberi kapcsolatok hálózatát vizsgáljuk, pusztán a fokszám változatosságával kapcsolatos intuícióinkra hagyatkozva is könnyen beláthatjuk, hogy a fokszámeloszlás nem követheti a Poisson-eloszlás alakját.

Ha ugyanis feltesszük, hogy a közvetlen ismerősök számának eloszlása úgy viselkedik, mint egy véletlen gráfban, és ehhez hozzávesszük azt a szociológiában ismert tényt, hogy egy személynek átlagosan $\langle k \rangle = 1000$ ismerőse van, akkor a 24. egyenlet alapján azt kapjuk, hogy a szórás ezen átlagérték körül $\sigma_{\mathrm{ER}} = \sqrt{\langle k \rangle} \simeq 31{,}62$. Ez azt jelenti, hogy egy tipikus egyén ismerőseinek száma $\langle k \rangle \pm \sigma$ körüli érték, vagyis 968 és 1032 közé esik, s ez egy meglehetősen szűk tartomány. Azt kaptuk tehát, hogy ha a fokszámeloszlás a klaszszikus véletlen gráf által jósolt Poisson-eloszlást követi, akkor nincsenek nagyon népszerű egyének a hálózatban, és olyan személyt sem találunk, aki kevés, csak egy-két ismerőssel rendelkezik.

Ez a meglepő következtetés a véletlen hálózatok azon fontos tulajdonságából adódik, hogy egy nagy véletlen hálózatban a legtöbb csomópont fokszáma nagyon közel esik a $\langle k \rangle$ értékhez. Ez az előrejelzés nyilvánvalóan ellentmond a valóságnak. Rengeteg emberről lehet ugyanis biztosan tudni, hogy az átlagnál lényegesen több ismerőse van. Például az Egyesült Államok egykori elnökének, Franklin Delano Rooseveltnek az előjegyzési naplójában nagyjából 22 ezer olyan ember neve áll, akikkel ő személyesen találkozott.[17] Egy, a Facebookról készült tanulmány szerint is igen sok azoknak a száma, akiknek a rendszer által maximálisan megengedett 5000 ismerősük van.[18]

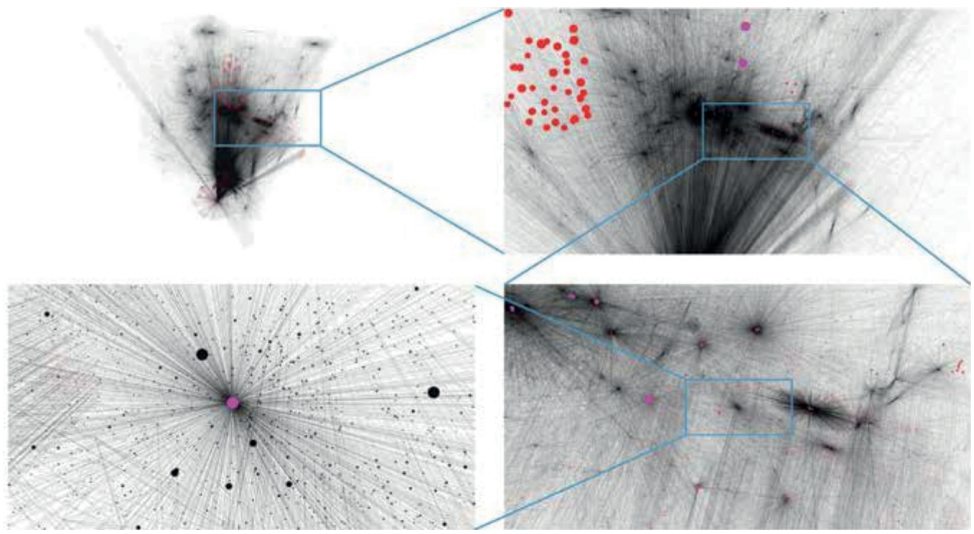
Nézzük meg ezek után, hogy konkrétan milyen alakot vesz fel a fokszámeloszlás valós hálózatokban. Történelmileg az első tanulmány, amely egy nagy méretű komplex

[16] Ha a Poisson-eloszlásból indulunk ki, akkor egyből a 24. egyenletben látható eredményre jutunk, ugyanis a Poisson-eloszlás esetén a várható érték és a szórásnégyzet megegyezik.

[17] Freeman–Thompson 1989.

[18] Backstrom et al. 2012.

hálózat fokszámeloszlását vizsgálta, Hawoong Jeonghoz és munkatársaihoz kötődik,[19] akik 1998-ban a körülbelül 300 ezer dokumentumból és 1,5 millió hivatkozásból álló nd.edu tartományt térképezték fel a világhálón kifejezetten abból a célból, hogy össze lehessen hasonlítani a web gráfjának és egy véletlen hálózat modelljének a tulajdonságait. 1998-ban csakugyan okkal lehetett azt gondolni, hogy a web jól közelíthető véletlen hálózatokkal.



24. ábra

Részletek a World Wide Web 1998-ban feltérképezett darabjáról

Megjegyzés: Az egymás után következő képek a hálózat egyre erősebb nagyítású helyi részleteit mutatják. Az első kép az összes (325 729) csomópontról ad képet a teljes adatállományról. Az 50 feletti kapcsolatot tartó csomópontokat pirossal jelöltük, az 500 feletti fokszámúakat pedig lilával. A kinagyított részletek néhány nagyon magas fokszámú csomópont (más néven, angolból átvett szóval hub) jelenlétét mutatják; ez a skálafüggetlen hálózatok velejárója.

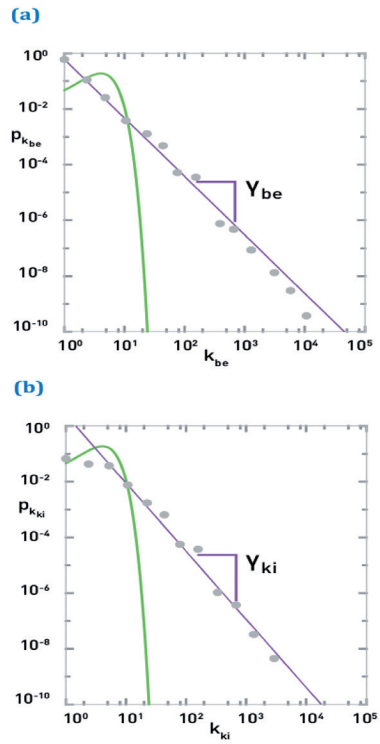
Forrás: BARABÁSI 2017

Minden dokumentum visszatükrözi készítőjének személyes és szakmai érdeklődését, ha egyénről van szó, ha szervezetről. Mivel az érdeklődési körök igen változatosak, azt lehetett gondolni, hogy ezekből a dokumentumokból véletlenszerűen kiválasztott dokumentumokra mutatnak a hivatkozások.

Első pillantásra ezt látszik igazolni a 24. ábra is: meglehetős véletlenszerűség van a web kapcsolati diagramja mögött. Ám ha közelebbről is megnézzük, feltűnhet néhány elgondolkodtató különbség a térkép és a véletlen gráf között. Mert véletlen hálózatban lényegében nem lehetnének az átlagtól jelentősen eltérő, magas fokszámú csomópontok (ezek angol neve: *hub*). A 24. ábrán viszont egyszerre van jelen sok alacsony fokszámú csomópont és néhány kivételesen magas fokszámú pont is.

A feltérképezett hálózat fokszámeloszlását a 25. ábrán láthatjuk.

[19] JEONG–ALBERT–BARABÁSI 1999.



25. ábra

A web fokszámeloszlása

Megjegyzés: A web hálózatának bejövő (a) és kimenő (b) fokszámeloszlása, amely mindkét tengelyen logaritmikus ábrázolásban (log–log ábrán) látható. Ha mindkét tengelyen logaritmikus a skála, akkor a hatványfüggvényt egyenes vonal jeleníti meg. A pontok a mért adatokat mutatják; a vonalak a mért pontokra illesztett hatványfüggvények, az egyiknek $\gamma_{be} = 2{,}1$ a fokszámkitevője, a másiké $\gamma_{ki} = 2{,}45$. A zöld vonal a web hálózatának átlagos $\langle k_{be} \rangle = \langle k_{ki} \rangle = 4{,}6$ fokszámához tartozó Poisson-eloszlást mutatja.

Forrás: BARABÁSI 2017

Ez jól tükrözi, hogy a Poisson-eloszlás rossz közelítés a web fokszámeloszlására. Log–log skálán viszont nagyjából egyenes vonalra esnek az adatpontok, s ez azt sejteti, hogy a web fokszámeloszlása jól becsülhető ezzel a képlettel:

$$p(k) \sim k^{-\gamma} \qquad 25.$$

A 25. egyenlet szerinti eloszlást *hatványfüggvény-eloszlásnak* nevezik. A γ szám itt a fokszámkitevő. Ha a 25. egyenlet mindkét oldalának logaritmusát vesszük, akkor a következőt kapjuk:

$$\ln p(k) \sim -\gamma \ln k \qquad 26.$$

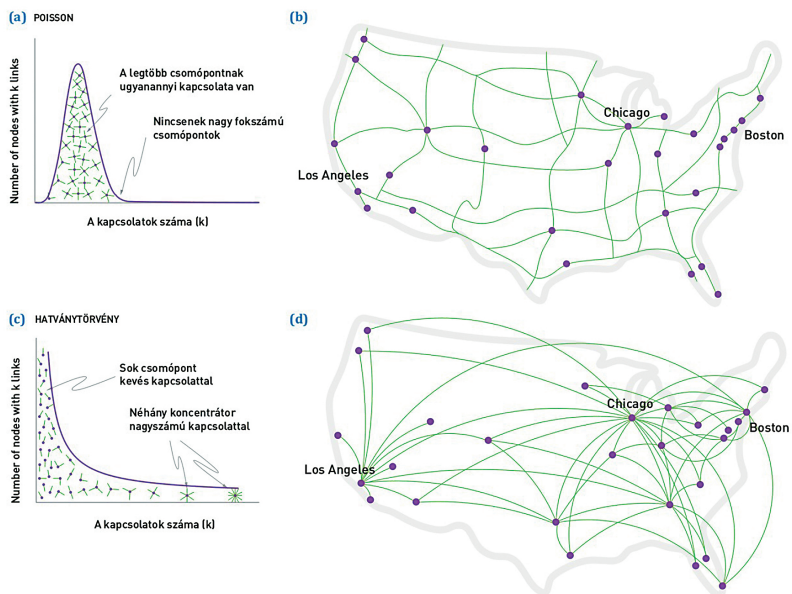
Ha a 26. egyenlet teljesül, akkor a $\ln p(k)$ lineárisan kell hogy függjön $\ln k$-tól, és az egyenes meredeksége a γ fokszámkitevő (25. ábra). A web irányított hálózat, ezért két fokszámeloszlást kell használnunk; külön a ki-fokszámokra vonatkozó $p(k_{\text{ki}})$ eloszlást [ahol $p(k_{\text{ki}})$ annak a valószínűsége, hogy egy véletlenszerűen kiválasztott csomópont pontosan k_{ki} számú másik webdokumentumra mutat], és a be-fokszámokra vonatkozó $p(k_{\text{be}})$ eloszlást.

A 25. ábra alapján a webgráfban mindkét fokszámeloszlás hatványfüggvényszerű viselkedést mutat, azaz

$$p(k_{\text{ki}}) \sim k^{-\gamma_{\text{ki}}} \quad , \quad p(k_{\text{be}}) \sim k^{-\gamma_{\text{be}}} \qquad 27.$$

ahol a két hatványkitevő a $\gamma_{\text{ki}} \simeq 2{,}45$ és $\gamma_{\text{be}} \simeq 2{,}1$ értékeket veszi fel. Az ilyen, hatványszerűen viselkedő fokszámeloszlással rendelkező hálózatokat hívjuk skálafüggetleneknek.

A klasszikus véletlen gráfra jellemző Poisson-eloszlás és a hatványszerű viselkedést mutató skálafüggetlen fokszámeloszlás közti legfontosabb eltéréseket próbálja szemléltetni a 26. ábra.



26. ábra

Véletlen és skálafüggetlen hálózatok

Megjegyzés: a) A véletlen hálózatok fokszámeloszlása Poisson-eloszlást követ, ezért a csomópontok többségének hasonló nagyságú a fokszáma, és a hálózatban nincsenek sok kapcsolatot fenntartó hubok. b) A véletlen hálózat hasonlít valamelyest az amerikai autópályarendszerhez; amelyben nincsenek több száz autópályás városok. c) A hatványfüggvény jellemezte hálózatokban a pontok döntő többségének csak kevés kapcsolata van. A nagyszámú „kis" csomópontot néhány kiemelkedően nagy fokszámú pont tartja össze. d) A skálafüggetlen hálózat úgy fest, mint a légi közlekedés hálózata, amelyben a legtöbb repülőtér kicsi, és csak néhány járat használja. De van néhány nagyon nagy repülőtér is – olyan, mint Chicago vagy Los Angeles –, és fontos középpontokként ezek kötik össze egymással a kisebbeket.

Forrás: BARABÁSI 2017

Ezen látható, hogy a Poisson-eloszlás és a skálafüggetlen fokszámeloszlás nagyon más alakúak. Véletlen hálózatokban egymáshoz hasonló a csomópontok fokszáma, a hubok tehát hiányoznak. A skálafüggetlen hálózatokban a hubok nem csupán lehetségesek, de várhatók is. Ezenfelül minél több csomópontból áll a skálafüggetlen hálózat, annál nagyobbak benne a hubok. Ismeretes, hogy a hubok mérete polinomiálisan nő a hálózat méretével, ezért igen nagyra megnőhetnek. A véletlen hálózatokban viszont csak a hálózat méretének (N-nek) a logaritmusával arányosan nő a legnagyobb fokszám, vagy még lassabban; emiatt még a nagyon nagy véletlen hálózatban sem jelennek meg hubok.

A skálafüggetlenség matematikai leírása

Ha definiálni szeretnénk a skálafüggetlenség fogalmát, akkor az előző alfejezet leírásánál kicsit precízebb, de még mindig viszonylag megengedő definíció az, hogy egy hálózat skálafüggetlennek tekinthető akkor, ha a fokszámeloszlása viszonylag széles fokszámtartományon hatványszerűen lecsengő viselkedést mutat. A 25. ábra b) képén például igaz a hatványfüggvényszerű lecsengés nagyjából 4 nagyságrenden keresztül (ami egy igen széles tartománynak minősül), de nem igaz a kis fokszámok tartományán. Ennek ellenére természetesen a web hálózatát továbbra is skálafüggetlennek tekinthetjük.

A skálafüggetlenségnek egy sokkal szigorúbb definíciója az, hogy a hálózatban lehetséges fokszámok tartományát egy $[k_{\min}, \infty]$ tartományon jelöljük ki (ahol $k_{\min} > 0$, illetve a legtöbb esetben a természetes választás a $k_{\min} = 1$), és megköveteljük, hogy a $p(k)$ ezen a teljes tartományon hatványszerűen viselkedjen. Természetesen a valós hálózatok esetén szinte teljesen biztos, hogy találunk olyan tartományt, ahol $p(k)$ nem követi precízen az adott hatványfüggvényt, ezért az ennek a szigorúbb definíciónak megfelelő fokszámeloszlás csak hálózati modellekben fordul elő, illetve különféle, hálózatelmélettel kapcsolatos analitikus számításoknál bukkan fel. Ezt a szigorúbb definíciót egyenletek formájában is megadhatjuk. A fokszám teljes értelmezési tartományára kiterjedő hatványszerű viselkedés azt jelenti, hogy

$$p(k) = A \cdot k^{-\gamma} \tag{28.}$$

ahol A egy konstans, amelynek értékét úgy kell megválasztani, hogy $p(k)$ normált legyen, azaz a

$$\sum_{k=k_{\min}}^{\infty} p(k) = \sum_{k=k_{\min}}^{\infty} Ak^{-\gamma} = 1 \tag{29.}$$

egyenlőség teljesüljön. Ez alapján $A = [\Sigma_k k^{-\gamma}]{-}1$, amelynél a zárójelen belül szereplő összeg a $k_{\min} = 1$ esetben a matematikában ismert Riemann-féle zeta-függvénybe megy át, amely a

$$\zeta(\gamma) = \sum_{k=1}^{\infty} k^{-\gamma} \tag{30.}$$

alakban adható meg. Ez alapján egy skálafüggetlen hálózat fokszámeloszlása a

$$p(k) = \frac{k^{-\gamma}}{\zeta(\gamma)} \qquad 31.$$

formában írható le.

Bizonyos analitikus levezetéseknél célravezető, ha a fokszámot (amely alapvetően diszkrét értékeket vehet fel) *folytonos változóként* kezeljük. Ez természetesen egy közelítés, viszont sok esetben nagyban leegyszerűsítheti a számolásokat. Folytonos esetben a fokszámeloszlás normáltságát összegzés helyett integrálással lehet felírni:

$$\int_{k_{\min}}^{\infty} p(k)\mathrm{d}k = \int_{k_{\min}}^{\infty} A \cdot k^{-\gamma}\mathrm{d}k = 1 \qquad 32.$$

ahol rögtön be is helyettesítettük a 28. egyenlet alakját (amely általános jellegénél fogva mind diszkrét, mind folytonos esetben alkalmazható). Az integrál könnyen elvégezhető, aminek segítségével egy egyszerű kifejezést kapunk az *A* konstansra:

$$\int_{k_{\min}}^{\infty} A \cdot k^{-\gamma}\mathrm{d}k = \left[\frac{A \cdot k^{1-\gamma}}{1-\gamma}\right]_{k_{\min}}^{\infty} = \frac{A \cdot k^{1-\lambda}}{\gamma - 1}, \rightarrow A = \frac{\gamma - 1}{k_{\min}^{1-\gamma}}, \qquad 33.$$

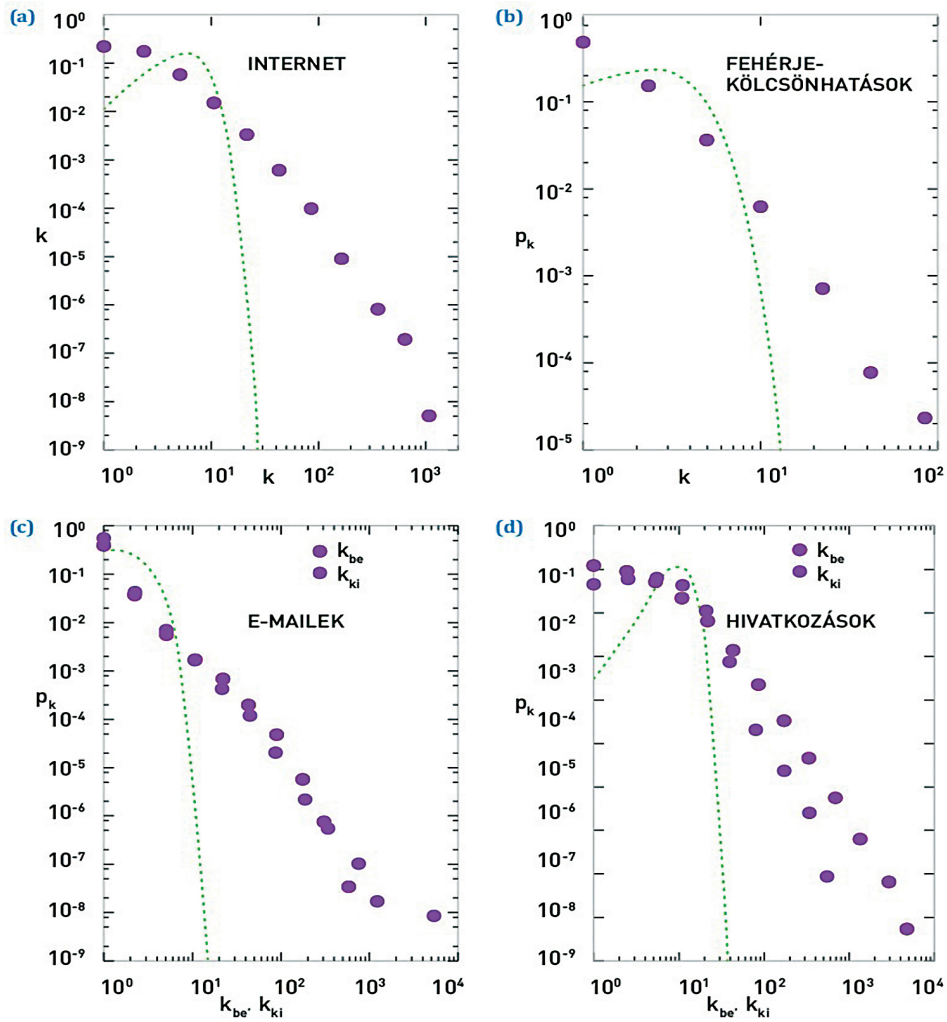
amit visszahelyettesítve a következő eredményre jutunk a folytonos skálafüggetlen fokszámeloszlás alakjára vonatkozóan:[20]

$$p(k) = \frac{(\gamma - 1)k^{-\gamma}}{k_{\min}^{1-\gamma}} \qquad 34.$$

A skálafüggetlenség univerzális

Az elmúlt közel két évtizedben sok tudományos, technológiai és társadalmi szempontból fontos valós hálózatról bizonyosodott be a skálafüggetlenség.

[20] Ez az eloszlási alak szélesebb körben a Pareto-eloszlás néven ismert, amelyet először Vilfredo Federico Damaso Pareto olasz közgazdász és politológus javasolt a társadalomban tapasztalható vagyoneloszlás jellemzésére.



27. ábra

A valós hálózatok többnyire skálafüggetlenek

Megjegyzés: A mintaként használt hálózatok közül négynek a fokszámeloszlása. Az ábrákon a valóságos hálózatot jellemző *p(k)* mellett zöld pontozott vonallal feltüntetjük a Poisson-eloszlást is, szemléltetve ezzel, hogy a klasszikus véletlengráf-modell nem ad magyarázatot a megfigyelt *p(k)*-ra. Ha irányított hálózatról van szó, akkor a ki- és a be-fokszámot külön ábrázoljuk. a) Az internet. b) A fehérje-kölcsönhatási hálózat. c) Az e-mail-hálózat. d) A tudományos hivatkozások.

Forrás: BARABÁSI 2017

Ezt szemlélteti a 27. ábra egy infrastrukturális hálózattal (az internettel), egy biológiai hálózattal (a fehérjék közötti kölcsönhatások hálózatával), egy kommunikációs hálózattal (az e-mailek hálózatával) és egy, a tudományos kommunikációt jellemző hálózattal

(a tudományos hivatkozás hálózatával). A fokszámeloszlás mindegyikben jócskán eltér a Poisson-eloszlástól, s jobban közelíthető hatványfüggvénnyel. A skálafüggetlen rendszerek szembetűnően sokfélék: a web egy alig két évtizedes, emberi kéz alkotta hálózat, a fehérjék közötti kölcsönhatások hálózata pedig négymilliárd éves evolúció eredménye. Bizonyos hálózatokban a csomópontok molekulák, másokban számítógépek. Ez a változatosság késztet bennünket arra, hogy a *skálafüggetlenséget univerzális* hálózati jellemzőnek tekintsük.

A kutató szemszögéből sarkalatos kérdés a következő: honnan tudjuk, hogy egy hálózat skálafüggetlen? Egyfelől elég egy gyors pillantás a fokszámeloszlásra, s az nyomban megmutatja, hogy a hálózat skálafüggetlen-e: skálafüggetlen hálózatban a legkisebb és a legnagyobb csomópont fokszáma meglehetősen eltér egymástól. Véletlen hálózatokban viszont összevethető a csomópontok fokszáma. A skálafüggetlen hálózatok fokszámeloszlásának ábrázolásakor általában célszerű a grafikon mindkét tengelyén logaritmikus skálát használni; az ilyen jellegű grafikonokat szokták „dupla logaritmikus" vagy „log-log" grafikonnak hívni. Azért célszerű ez az ábrázolásmód, mert ha a csomópontok fokszáma nagyon erősen eltér egymástól, akkor nem lehet mindet egy lineáris ábrán megjeleníteni. Ahhoz, hogy jól láthatóan megkapjuk a jegyzetben bemutatott fokszámeloszlásokat, logaritmikus „dobozolást" (binning) használtunk; ezzel elérhető, hogy minden adatpont statisztikailag megfelelő legyen.

Mivel a fokszámeloszlás kitevőjének értéke fontos a különböző hálózati tulajdonságok előrejelzésében, ha ezt szeretnénk kinyerni, akkor egy egyszerű megközelítésben megkaphatjuk függvényillesztéssel, de természetesen vannak erre precízebb statisztikai módszerek is. Fontos megjegyezni, hogy a valóságos hálózatokban tapasztalt fokszámeloszlás nemegyszer eltér a tiszta hatványfüggvénytől. Ezek az eltérések fakadhatnak az adatok fogyatékosságából, torzító hatású adatgyűjtésből, de fontos információt is tartalmazhatnak a hálózatokat kialakító folyamatokról.

A skálafüggetlenség jelentése és következményei

A „skálafüggetlen" kifejezés a statisztikus fizika egyik ágából, a fázisátalakulások elméletéből származik, ennek a tudományterületnek a kutatói az 1960-as, 1970-es években átfogóan vizsgálták a hatványfüggvény-szerűen viselkedő eloszlásokat. Hogy jobban megértsük a skálafüggetlenséget, jobban meg kell ismernünk a fokszámeloszlás momentumait.

A fokszámeloszlás n-edik momentumának definíciója a következő:

$$\left\langle k^n \right\rangle = \sum_k k^n p(k) = \int_{k_{\min}}^{\infty} k^n p(k) \mathrm{d}k \qquad 35.$$

ahol mind a diszkrét, mind a folytonos eloszlást használó alakot feltüntettük. Az alacsonyabb momentumoknak fontos jelentésük van:

- $n = 1$: az első momentum az átlagos fokszám, $\langle k \rangle$.
- $n = 2$: a második momentum, $\langle k^2 \rangle$ jól jön a szórásnégyzet (variancia) kiszámításában, ahogy azt már a 22. egyenletnél láthattuk, $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$. A szórásnégyzet azt méri, hogy a fokszámok mennyire térnek el egymástól, azaz átlagosan mekkorák a fluktuációk az átlagérték körül.

- $n = 3$: a harmadik momentum $\langle k^3 \rangle$ határozza meg az eloszlás ferdeségét, azaz megmutatja, hogy a $p(k)$ mennyire szimmetrikus a $\langle k \rangle$ érték körül.

Skálafüggetlen hálózatokban a fokszámeloszlás n-edik momentuma (véges számú pontot, és emiatt egy véges $k_{\max}$ maximális fokszámot feltételezve) a következő alakban írható fel:

$$\left\langle k^n \right\rangle = \int_{k_{\min}}^{k_{\max}} k^n \frac{(\gamma-1)k^{-\gamma}}{k_{\min}^{1-\gamma}} \mathrm{d}k = \frac{(\gamma-1)\left[k_{\max}^{n+1-\gamma} - k_{\min}^{n+1-\gamma}\right]}{(n+1-\gamma)k_{\min}^{1-\gamma}} \qquad 36.$$

ahol az egyszerűség kedvéért a folytonos alakját használtuk a fokszámeloszlásnak. Egy további egyszerűsítés gyanánt tegyük fel, hogy $k_{\min} = 1$, ami amúgy egy nagyon természetes alsó határ. Ilyenkor a 36. egyenlet a

$$\left\langle k^n \right\rangle = \frac{(\gamma-1)\left[k_{\max}^{n+1-\gamma} - 1\right]}{n+1-\gamma} \qquad 37.$$

képletbe megy át. Ha megvizsgáljuk ennek a kifejezésnek a viselkedését nagy hálózatok esetén, akkor mivel skálafüggetlen hálózatokban N növekedtével $k_{\max}$ szintén (polinomiálisan) nő, az $N \to \infty$ (és ezzel együtt $k_{\max} \to \infty$) határesetben két dolog történhet:

- Amennyiben $n + 1 - \gamma < 0$, úgy $k_{\max}$ egy negatív hatványon szerepel a 37. egyenletben, ezért az n-edik momentum, $\langle k^n \rangle$ egy jól definiált véges értékhez tart.
- Ha viszont $n + 1 - \gamma > 0$, akkor $k_{\max}$ kitevője már pozitív, ennélfogva a 37. egyenlet egy minden határon túl növekvő, $\langle k^n \rangle \to \infty$ eredményt ad az n-edik momentumra.

Ebből az okfejtésből az következik, hogy ha például a skálafüggetlen hálózat γ exponense $\gamma < 3$, akkor már az $n = 2$ momentum is divergál, azaz az $N \to \infty$ esetben a fokszámeloszlás szórása is minden határon túl nő. Ezt a divergenciát látva megérthetjük, miből ered a „skálafüggetlen" jelző. A σ szórás egyik szemléletes jelentése az, hogy ha véletlenszerűen választunk egy pontot a hálózatból, akkor annak fokszáma jó eséllyel a

$$k = \langle k \rangle \pm \sigma \qquad 38.$$

tartományba esik. A $\langle k \rangle$ és a σ viszont egészen más nagyságrendű a véletlen és a skálafüggetlen hálózatokban:

- *A véletlen hálózatokban van jellemző méretskála:* Poisson-eloszlás szerinti fokszámeloszlású véletlen hálózatokban $\sigma = \langle k \rangle^{1/2}$, s ez mindig kisebb, mint $\langle k \rangle$. Ezért a hálózat csomópontjainak fokszáma a $k = \langle k \rangle \pm \langle k \rangle^{1/2}$ tartományba esik. Más szóval, a véletlen hálózatok csomópontjainak fokszáma hasonló, és az átlagos fokszám, a $\langle k \rangle$ a véletlen hálózat „méretskálája".
- *A skálafüggetlen hálózatoknak nincsen jellemző méretskálájuk:* Egy olyan hálózatban, amelyben a fokszámeloszlást $\gamma < 3$ kitevőjű hatványfüggvény adja meg, az első momentum véges, a második azonban már végtelen. A $\langle k^2 \rangle$ (és a σ) nagy N értéken tapasztalható divergenciája (végtelenbe tartása) azt mutatja, hogy az átlag körüli ingadozás tetszőleges nagyságú lehet. Ez annyit tesz, hogy ha véletlenszerűen

kiválasztunk egy csomópontot, akkor nem tudni, mire számíthatunk: a kiválasztott csomópont fokszáma lehet kicsi és hatalmas is.

Ezért a $\gamma < 3$ hálózatokban nincs értelmezhető belső skála; ezek a hálózatok skálafüggetlenek.

5. táblázat

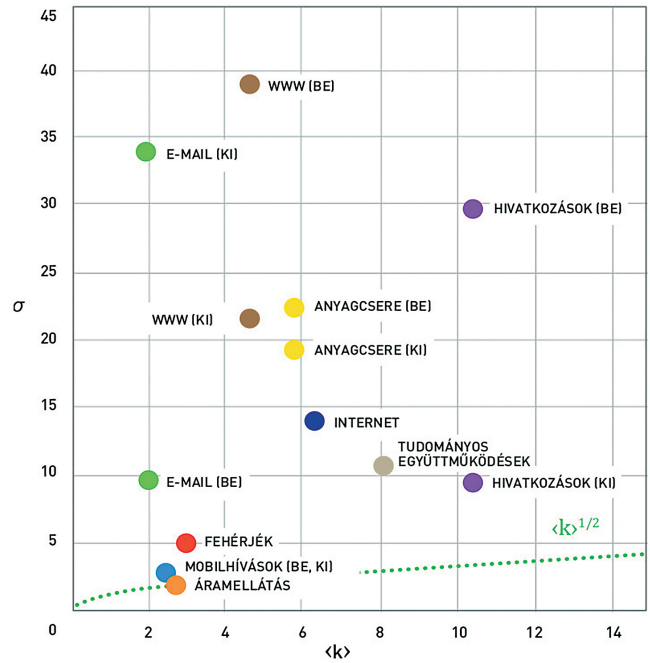
Fokszámingadozás a mintaként használt tíz hálózatban

Hálózat neve	N	L	$\langle k \rangle$	$\langle k_{be}^2 \rangle$	$\langle k_{ki}^2 \rangle$	$\langle k^2 \rangle$	γ_{be}	γ_{ki}	γ
Internet	192 244	609 066	6,34	–	–	240.1	–	–	3,42
WWW	325 729	1 497 134	4.60	1546,0	482,4	–	2,00	2,31	–
Áramellátás	4 941	6 594	2,67	–	–	10,3	–	–	Exp.
Mobilhívások	36 595	91 826	2,51	12,0	11,7	–	4,69	5,01	–
E-mail	57 194	103 731	1,81	94,7	1163,9	–	3,43	2,03	–
Tudományos együttműködések	23 133	93 439	8,08	–	–	178,2	–	–	3,35
Színészek hálózata	702 388	29 397 908	83,71	–	–	47 353,7	–	–	2,12
Hivatkozási hálózat	449 673	4 689 479	10,43	971,5	198,8	–	3,03	4,00	–
E. coli anyagcseréje	1 039	5 802	5,58	535,7	396,7	–	2,43	2,90	–
Élesztő fehérjéinek kölcsönhatásai	2 018	2 930	290	–	–	32,3	–	–	2,89

Megjegyzés: Az átlag körüli fluktuációk nagysága a 2. momentumtól, $\langle k^2 \rangle$-től függ; ezt a 4–6. oszlopokban tüntettük fel (különválasztva a ki- és a be-fokszámok eloszlását, valamint az irányítatlan esetet). Az, hogy $\langle k^2 \rangle$ nagy értéket vesz-e fel, szoros kapcsolatban áll a γ exponenssel, amely a 7–10. oszlopokban szerepel.

Forrás: BARABÁSI 2017

A fentieket például a webminta be-fokszámaira alkalmazva azt kapjuk, hogy az átlagos fokszám $\langle k \rangle = 4{,}6$ és $\gamma = 2{,}1$, ezért a második momentum divergál, vagyis egy véletlenszerűen választott webdokumentum várható be-fokszáma az $N \to \infty$ határéresetben $\langle k \rangle = 4{,}6 \pm \infty$. Vagyis egy véletlenszerűen választott webdokumentum fokszáma könnyen lehet egy vagy kettő, hiszen a csomópontok 74,02%-ának a be-fokszáma $\langle k \rangle$ alatt van. De lehet persze egy több százmillió kapcsolat jellemezte csomópont is – olyan, mint a google.com vagy a facebook.com.



28. ábra

A mintaként használt hálózatokban a fokszám szórása

Megjegyzés: A véletlen hálózatok $\sigma = \langle k \rangle^{1/2}$ szórását a zöld szaggatott vonal jelzi az ábrán. A színészek hálózatát (amelyben igen nagy a $\langle k \rangle$ és a σ) a jobb áttekinthetőség kedvéért lehagytuk.

Forrás: BARABÁSI 2017

Matematikailag pontos értelemben $\langle k^2 \rangle$ csak az $N \to \infty$ határesetben divergál. A divergencia azonban már egy véges hálózatra is rányomja a bélyegét. Az 5. táblázat és a 28. ábra ezt mutatja meg a jegyzetben mintaként használt valós hálózatokon. Ezekben a hálózatokban a $\langle k^2 \rangle$ és a σ általában jóval nagyobb, mint $\langle k \rangle$, vagyis a csomópontok fokszáma között nagy az eltérés. Például egy, a webmintából véletlenszerűen választott csomópont be-fokszáma $k_{\text{be}} = 4{,}6 \pm 39$, azaz a szórás majdnem tízszer akkora, mint maga az átlagérték. Ebből megint az következik, hogy az átlag értéke sokat nem mond.

Összefoglalva: a skálafüggetlenség annyit tesz, hogy nincs belső skála – azért nincs, mert ugyanabban a hálózatban nagyon különböző a csomópontok fokszáma. A skálafüggetlen hálózatokat ez a tulajdonság különbözteti meg egyfelől a rácsoktól – azokban minden csomópontnak ugyanaz a fokszáma ($\sigma = 0$) –, másfelől a klasszikus véletlen gráftól, amelyben a fokszámok szűk tartományba ($\sigma = \langle k \rangle^{1/2}$) esnek.

A skálafüggetlen fokszáméloszlásnak nemcsak a $p(k)$ momentumaira nézve érdekes matematikai következményei vannak, hanem a hálózat egyéb tulajdonságait is érdemben befolyásolhatja. Itt ezeket az effektusokat csak egy rövid, tényszerű felsorolás erejéig ismertetjük.

- *„Ultra kis világ" tulajdonság.* A skálafüggetlen hálózatokban az átlagos távolság $\gamma<3$ esetén még az Erdős–Rényi-gráfban tapasztaltnál is lassabban nő a rendszerméretttel:[21]

$$\langle d \rangle \sim \begin{cases} \text{konst.} & \gamma \le 2 \\ \dfrac{\ln\ln N}{\ln(\gamma-1)} & 2 < \gamma \le 3 \end{cases} \text{ Ultra kis világ}$$
$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad 39.$$
$$\begin{cases} \dfrac{\ln N}{\ln\ln N} & \gamma = 3 \\ \ln N & \gamma > 3 \end{cases} \text{ Kis világ}$$

- Ezt a viselkedést szokás „ultra kis világ" tulajdonságnak nevezni, amelynek intuitív magyarázata az, hogy a hubok kiemelten nagy szerepet játszanak a rövid elérési útvonalak biztosításában, hiszen például az összes szomszédjuk között rögtön egy maximum két lépésből álló utat nyújtanak. Minél kisebb a γ, annál dominánsabbak a hubok, és ezzel párhuzamosan annál kisebb az átlagos legrövidebb úthossz.
- *Robusztusság.* A skálafüggetlen hálózatok robusztusabbak véletlen csúcseltávolítási folyamatokkal szemben, mint a klasszikus véletlen gráf.[22] Amint korábban említettük, egy hálózat legalapvetőbb funkciója az, hogy közvetlen vagy közvetett elérést biztosít a pontjai között egymáshoz, és amennyiben a rendszer tartalmaz egy óriás összefüggő komponenst, úgy vélhetően ezt a feladatát el is látja. Azonban nagyon sok valós hálózat esetén realisztikus feltételezni azt, hogy a csomópontok egy része időről időre eltűnhet a hálózatból, például meghibásodik egy router a számítógép-hálózatban, vagy egy betegség miatt kiesik egy fehérje a fehérje-kölcsönhatási hálózatból. A kérdés az, hogy az ilyen, véletlenszerűnek tekinthető események során a csomópontok mekkora hányadának eltávolítása esetén bomlik fel az óriás komponens sok kicsi, egymástól izolált komponensre? Az ilyen jellegű szimulációk és analitikus számolások azt mutatják, hogy ez a kritikus, eltávolított ponthányad szignifikánsan nagyobb skálafüggetlen hálózatokban (megközelítve a teljes csúcshányadot) egy klasszikus véletlen gráfhoz képest (amelyben ez az érték tipikusan jóval kisebb a teljes hányadnál). Ez alapján egy skálafüggetlen hálózat sokkal jobban ellenáll a véletlenszerű „meghibásodásoknak" megfelelő, véletlenszerű csúcseltávolítási folyamatoknak, mint az Erdős–Rényi-féle véletlen gráf (azaz ebből a szempontból sokkal robusztusabb). Ennek szemléletes magyarázata az, hogy a hubok nagyon erősen összetartják a rendszert, és amíg ezek megmaradnak, addig jó eséllyel az óriás komponens sem esik szét. Ugyanakkor a hubok relatíve kevesen vannak a csomópontok összességéhez képest, ezért véletlenszerű választással nehéz őket „eltalálni".
- *Érzékenység célzott támadással szemben.* Az imént tárgyalt extrém robusztusságnak van egy árnyoldala is. Ha nem véletlenszerű sorrendben távolítjuk el a csomópontokat, hanem a fokszám szerinti sorrendben, akkor a skálafüggetlen hálózatok

[21] BOLLOBÁS–RIORDAN 2004; COHEN–HAVLIN 2003.
[22] ALBERT–JEONG–BARABÁSI 2000.

hirtelen nagyon törékennyé válnak, és már sokkal kisebb csúcshányad eltávolítása után széteshetnek, mint ha ugyanezt egy Erdős–Rényi-hálózatban tesszük.[23] Ez a jelenség megint a hubok fontosságára világít rá, hiszen egy skálafüggetlen hálózat esetén ennél az eltávolítási folyamatnál a hubokat vesszük ki először a hálózatból. Az Erdős–Rényi-gráf esetén viszont, mivel nagyon homogén a fokszámeloszlás, szinte mindegy, hogy a fokszám szerinti sorrendben távolítjuk el a csomópontokat, vagy teljesen véletlenszerűen.

- *Anomális terjedési jelenségek.* A különféle terjedési folyamatok (például járvány- vagy információterjedés stb.) hálózaton történő modellezése során általában be szokás vezetni egy úgynevezett terjedési rátát, amely megadja, hogy a fertőzés mekkora valószínűséggel terjed tovább egy már fertőzött pontból egy még egészséges közvetlen szomszédja felé egységnyi idő alatt. A szimulációk és analitikus számítások alapján egy homogén hálózatban, mint például a klasszikus véletlen gráf, a terjedési rátának el kell érnie egy kritikus értéket ahhoz, hogy a fertőzés el tudjon terjedni a hálózatban, ellenkező esetben a betegség rövid úton kihal, és mindenki egészséges lesz. Ezzel szemben skálafüggetlen hálózatokban érdekes módon ez a kritikus terjedési ráta nullához tart,[24] ami azt jelenti, hogy olyan gyenge fertőzések is el tudnak terjedni a rendszerben, amelyeknek egy Erdős–Rényi-gráfban erre semmi esélyük sem lenne.

Fokszám-korrelációk

Közismert, hogy a hollywoodi sztárok előszeretettel házasodnak egymás között, azaz egy hollywoodi sztár párja jó eséllyel egy másik hollywoodi sztár lesz. Bár ezt sokszor természetesnek vesszük, mégis, ha matematikai szempontból vizsgáljuk ennek esélyét, akkor meglepően kicsi valószínűséget kapunk eredményül. Az egyszerűség kedvéért tegyük fel, hogy egy sztár bárkivel randevúzhat a világon élő kb. százmillió személyből. Így annak az esélye, hogy választottja szintén rajta lesz a sztárok kb. ezres listáján, nagyjából 10^{-5}, ami elenyészően kevés (gyakorlatilag nulla) sztárpár-előfordulást jelentene. Hasonló jelenséget figyelhetünk meg politikai vezetők és fontos vállalatok vezérigazgatói esetén is, akik egyfelől az átlagosnál sokkal több kapcsolattal rendelkeznek (azaz a fokszám szempontjából hubként viselkednek), másrészt egymást is jó eséllyel ismerik (azaz a többi hubhoz is van kapcsolatuk).

Érdekes módon ez a jelenség, amely szerint a nagy fokszámú pontok előszeretettel kapcsolódnak a többi nagy fokszámú csúcshoz, nem minden hálózatban magától értetődő.

[23] Albert–Jeong–Barabási 2000.

[24] Pastor-Satorras–Vespignani 2001.

29. ábra

Az élesztő fehérje-kölcsönhatási hálózata

Megjegyzés: Minden csúcspont egy-egy fehérjének felel meg; két fehérje közt akkor van él, ha kísérleti bizonyíték van arra, hogy fizikailag összekapcsolódhatnak a sejtben. Kiemeltük a két legnagyobb, $k = 56$ és $k' = 13$ fokszámú középpontot. Mindkettő sok kis fokszámú csomóponthoz kapcsolódik, de egymással nem szomszédok.

Forrás: JEONG et al. 2001

A 29. ábrán például egy fehérje-kölcsönhatási hálózatot láthatunk, amelyben a hubok inkább kis fokszámú pontokhoz kapcsolódnak (egyfajta középpont-küllő szerkezetet alkotva), és egymáshoz nem. Ez különösen szembetűnő a 29. ábrán kiemelt két középpont esetében: szinte csak kis fokszámú fehérjékhez kapcsolódnak.

Egy rövid számítás nyomban igazolja, mennyire szokatlan is ez a szerkezet. Tegyük fel, hogy minden csomópont esetén megtartjuk az eredeti fokszámot, viszont teljesen véletlenszerűen választjuk ki, hogy melyik másik pontokkal kötjük össze. Ezt úgy a legkönnyebb

elképzelni, hogy az eredeti hálózatban minden kapcsolatot félbevágunk, átmenetileg izolálva a csúcsokat egymástól, majd a fél éleket teljesen véletlenszerűen kapcsoljuk össze ismét. Ezáltal létrehozunk egy olyan új hálózatot, amelyben minden pontnak annyi a fokszáma, mint az eredeti fehérje-kölcsönhatási hálózatban, viszont teljesen véletlenszerű, hogy melyik melyikkel szomszédos.

Vizsgáljuk meg, mennyi az esélye annak, hogy a hálózat ezen véletlenszerű másolatában egy k_i és egy k_j fokszámú csúcs között van kapcsolat. Amikor a fél élek újra összekapcsolásánál egy i-ből induló fél él kiválasztja, hogy melyik másik fél éllel alkot egy teljes kapcsolatot ismét, annak a valószínűsége, hogy a j-hez kötődő fél élekből választ, $k_j/2L$, hiszen összesen $2L$ fél él van a rendszerben. Mivel i összesen k_i-szer „próbálkozik", egy elég nagy hálózat esetén feltehetjük, hogy maximum egyszer sikerül összekapcsolódni j-vel, és így annak valószínűsége, hogy i és j szomszédos, egyszerűen

$$p_{k_i,k_j} = \frac{k_i k_j}{2L} \qquad 40.$$

A 40. egyenlet azt mutatja, hogy a sok más csomóponthoz kapcsolódó középpontok sokkal szívesebben kapcsolódnak egymáshoz, mint kis fokszámú csomópontokhoz. Vagyis, ha k_i és k_j nagy, akkor p_{k_i,k_j} is nagy. Következésképpen annak a valószínűsége, hogy a $k_i = 56$ és a $k_j = 13$ fokszámú középpont közvetlenül kapcsolódjon egymáshoz: $p_{k_i,k_j} = 0{,}16$, s ez 400-szor nagyobb, mint a $p_{1,2} = 0{,}0004$ (ez utóbbi annak valószínűsége, hogy egy 2 fokszámú pont és egy 1 fokszámú pont egymáshoz kapcsolódnak). A 29. ábrán még sincs közvetlen kapcsolat a középpontok között, viszont sok közvetlen kapcsolat figyelhető meg a kis fokszámú csomópontok között. A 29. ábrán kiemelt középpontok az összekapcsolódás helyett majdnem kizárólag 1 fokszámú csomópontokhoz kapcsolódnak. Önmagában ez nem meglepő: azt várnánk, hogy a $k = 56$ fokszámú középpont $(N-1)p1{,}56 \simeq 12$ darab $k = 1$ fokszámú csomóponthoz kapcsolódik. A probléma az, hogy ez a középpont 46 darab 1 fokszámú szomszédjához kapcsolódik, azaz a vártnál négyszerte többhöz.

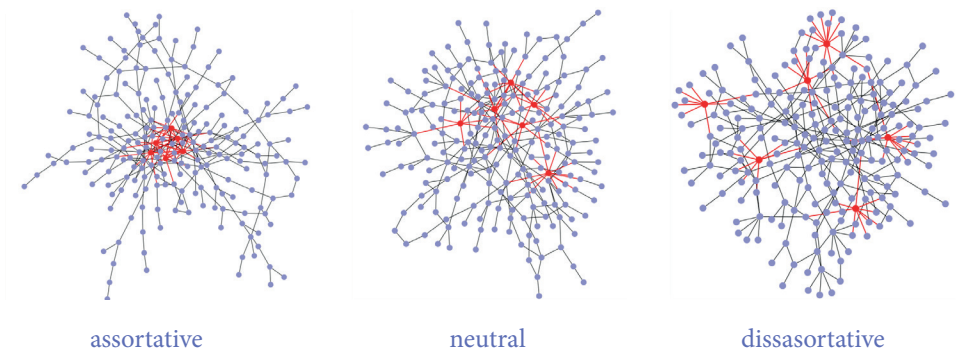
Összefoglalva: a szociális hálózatban úgy tűnik, a hubok szeretnek egymáshoz kapcsolódni, a fehérje-kölcsönhatási hálózatra viszont ennek éppen az ellenkezője igaz: a középpontok kerülik az egymáshoz kapcsolódást, ahelyett sok, kis fokszámú csomóponthoz kapcsolódnak. Bár veszélyes a két példa alapján általános alapelveket megfogalmazni, ennek a fejezetnek mégis az a célja, hogy megmutassa: ezekben a mintákban a valóságos hálózatok általános jellemzői mutatkoznak meg: egy fokszám-korrelációnak nevezett jelenségről tanúskodnak. Bemutatjuk a fokszám-korrelációk mérését, és azt, hogyan lehet megismerni, milyen hatásuk van a hálózati topológiára.

Asszortatív, neutrális és diszasszortatív hálózatok

Láttuk, hogy bizonyos hálózatokban a nagy fokszámú csúcsok nagyobb eséllyel kapcsolódnak egymáshoz, mint amit egy véletlenszerű kapcsolódás alapján feltételeznénk, míg más hálózatokban ezzel homlokegyenest ellenkezőleg: sokkal kisebb annak a valószínűsége, hogy közvetlen kapcsolatot találjunk két hub között, mint amit egy fokszámtól független

kapcsolódási mechanizmus esetén kapnánk. Alapvetően három osztályba szokás sorolni a hálózatokat ezen tulajdonság alapján:

- *Asszortatív hálózatok:* ezekben a hálózatokban a nagy fokszámú pontok előszeretettel kapcsolódnak más nagy fokszámú pontokhoz, és ezzel párhuzamosan a kis fokszámú pontok is inkább kis fokszámú pontokhoz szeretnek kapcsolódni. Ilyenkor a szomszédos csomópontok fokszámai között *pozitív korreláció* lép fel, hiszen egy nagy fokszámú pont közvetlen szomszédja jó eséllyel szintén nagy fokszámú, és vice versa, egy kis fokszámú pont közvetlen szomszédja szintén jó eséllyel kis fokszámú.
- *Neutrális hálózatok:* ezekben a hálózatokban a szomszédos csomópontok fokszáma függetlennek tekinthető egymástól, azaz a szomszédos pontok fokszáma között *gyakorlatilag nincs korreláció.* Ilyenkor a 40. egyenlet nagyjából jól írja le két adott k_i és k_j fokszámú csúcspár között a közvetlen kapcsolat létezésének valószínűségét.
- *Diszasszortatív hálózatok:* ezekben a hálózatokban a nagy fokszámú pontok előszeretettel kapcsolódnak kis fokszámú csúcsokhoz és fordítva, a kis fokszámú csúcsok is nagyobb eséllyel lesznek egy hub szomszédai, mint amit véletlenszerű kapcsolódás esetén feltételeznénk. Ilyenkor a szomszédos pontok fokszámai között *negatív a korreláció.*



30. ábra

Asszortatív, neutrális és diszasszortatív hálózatok

Megjegyzés: A fokszám-korreláció viselkedése alapján alapvetően háromfajta viselkedésről beszélhetünk: az asszortatív hálózatokban a hubok (pirossal kiemelve) szeretnek egymáshoz kapcsolódni, ezért előfordul, hogy egy „elit klubot" hoznak létre a hálózat centrumában; a neutrális hálózatban nem korrelált a szomszédok fokszáma; míg a diszasszortatív hálózatokban a hubok „taszítják" egymást, és általában nem hoznak létre közvetlen kapcsolatot egymással.

Forrás: BARABÁSI 2017

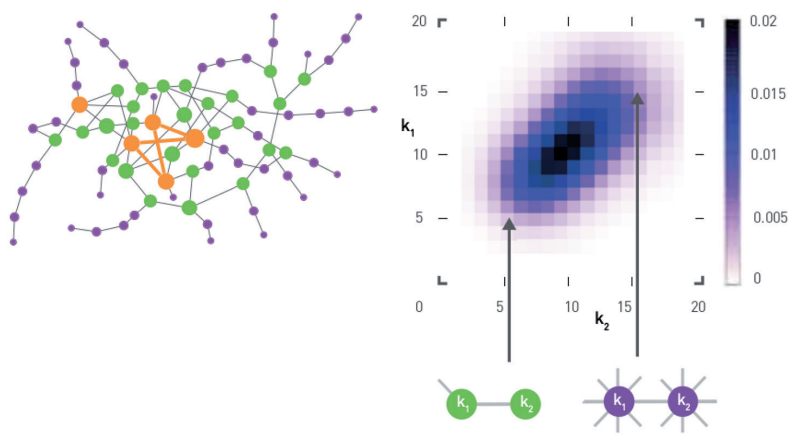
A 30. ábrán sematikus illusztrációt mutatunk mindhárom esetre. Látható, hogy az asszortatív hálózatokban a pozitív fokszám-korrelációk egyik szemléletes következménye lehet, hogy a nagy fokszámú csúcsok egyfajta „elit klubot" formálnak egymással, amely egy nagyon erős magot képezhet a hálózat centrumában. Ezzel szemben a diszasszortatív hálózatokban a hubok körül tipikusan egy-egy középpont-küllők szerkezet (vagy más néven csillag) alakul ki, hiszen a szomszédaik döntő többsége kis fokszámú (ami kizárja, hogy ezen szomszédok között pluszban még sok él létezhessen).

A fokszám-együttelőfordulási mátrix

A fokszám-korreláció mérésének egyik kézenfekvő módszere az, amikor végigmegyünk a kapcsolatok listáján, és minden kapcsolat esetén feljegyezzük az adott él végén látott fokszámok együtt-előfordulását. Ezt legegyszerűbb egy olyan négyzetes $E_{k,k'}$ mátrix segítségével végezni, amely annyi sorból és oszlopból áll, mint ahányféle fokszám előfordul a hálózatban, és ezáltal minden mátrixelem megfeleltethető egy-egy lehetséges fokszámpárnak:

- A mátrixelemeket a procedúra elején nullára állítjuk be, azaz mielőtt végigmennénk az éleken, $E_{k,k'} = 0$ minden lehetséges k és k' értékére.
- Majd sorban megvizsgáljuk a kapcsolatokat az élek végén található két pont fokszámának szempontjából, amelyeket jelöljünk k-val és k'-vel. Ha az adott él esetén $k \neq k'$, akkor mind az $E_{k,k'}$, mind az $E_{k',k}$ értéket 1-gyel növeljük, hiszen az adott kapcsolat azt jelenti, hogy egy k fokszámú csúcs szomszédos egy k' fokszámú csúccsal, és fordítva. Amennyiben az adott kapcsolat azonos k fokszámú pontokat köt össze, azaz $k = k'$, akkor a mátrix főátlójában elhelyezkedő $E_{k,k}$ mátrixelem értékét 2-vel növeljük. Ez azért fontos, mert így biztosított, hogy minden figyelembe vett él összesen 2-vel növeli a mátrixelemek összegét.

Ha végeztünk a kapcsolatok listájával, és „feltöltöttük" az $E_{k,k'}$ mátrixot, érdemes minden mátrixelemet elosztani $2L$-el, ezzel egy olyan új $e_{k,k'} = E_{k,k'}/2L$ mátrixot kapunk, amelyben az elemek összege garantáltan 1, azaz tekinthetünk úgy rájuk, mint valószínűségekre. Ebben az új mátrixban adott k és k' esetén az $e_{k,k'}$ mátrixelem annak valószínűségét adja meg, hogy ha véletlenszerűen választunk a hálózatban egy kapcsolatot, akkor az épp egy k és egy k' fokszámú csúcs között fog húzódni.



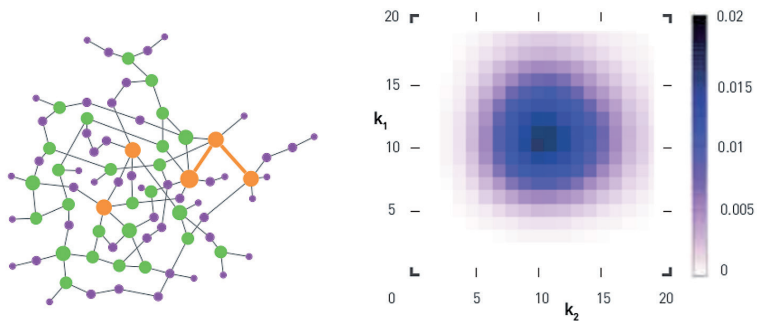
31. ábra

Az $e_{k,k'}$ mátrix asszortatív hálózatokban

Megjegyzés: Bal oldalon egy szemléltető ábra, jobb oldalon pedig a mátrix, amelyben az elemek nagyságát színkódolással jelezzük.

Forrás: Barabási 2017

Az $e_{k,k'}$ mátrix viselkedését szemlélteti a 31. ábra asszortatív esetben. Mivel egy asszortatív hálózatban a hasonló fokszámú csúcsok között várunk nagyobb élvalószínűséget, a mátrixban az azonos fokszámú fokszámpároknak megfelelő átló mentén lesznek relatíve nagy értékek, ettől az átlótól távol pedig kis értékek.



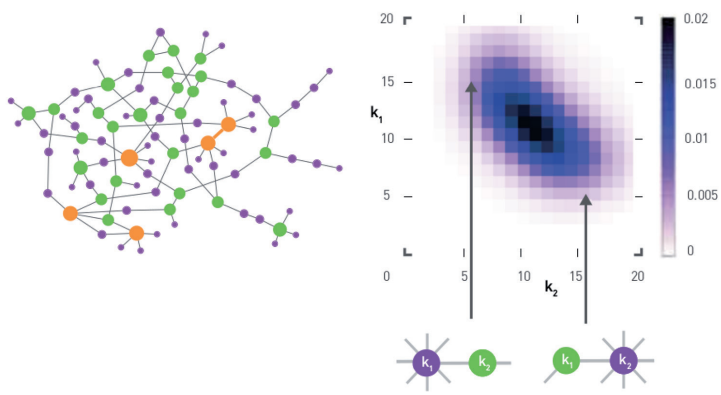
32. ábra

Az $e_{k,k'}$ mátrix neutrális hálózatokban

Megjegyzés: A mátrixban (amely a bal oldalon látható) ezúttal nincs kitüntetve a főátló, mivel a szomszédos csúcsok fokszáma gyakorlatilag függetlennek tekinthető.

Forrás: BARABÁSI 2017

Ezzel szemben egy neutrális hálózat esetén már nincs kitüntetett szerepe a főátlónak, ahogy azt a 32. ábrán láthatjuk. Ilyenkor a relatíve magasabb értékű mátrixelemek az előzőleg látott elnyújtott ellipszoid helyett tipikusan inkább egy kevésbé strukturált, körszerű alakot formálnak.



33. ábra

Az $e_{k,k'}$ mátrix diszasszortatív hálózatokban

Megjegyzés: Mivel az ilyen hálózatokban a hubok kis fokszámú pontokkal szeretnek összekapcsolódni és fordítva, a 31. ábrával ellentétes irányultságú alakot vesz fel a nagyobb mátrixelemek halmaza.

Forrás: BARABÁSI 2017

Végezetül a 31. ábrával ellentétes viselkedést láthatunk a 33. ábrán, amely az $e_{k,k'}$ tipikus alakját szemlélteti diszasszortatív hálózatok esetén. Látható, hogy a nagyobb értéket képviselő mátrixelemek itt is egy nyújtott ellipszoid alakot formálnak, azonban ennek irányultsága a mátrix másik átlóját követi.

Az $e_{k,k'}$ mátrix kétségtelen előnye, hogy részletesen tartalmazza az összes olyan információt, amelyekre szükségünk lehet a fokszám-korrelációk kiértékelésénél, hiszen minden lehetséges fokszámpárra megtaláljuk benne a hozzá tartozó előfordulási valószínűséget. Azonban közvetlen használatának van néhány hátránya is:

- Előfordulnak gyengén asszortatív vagy diszasszortatív hálózatok, amelyek esetén a mátrix vizuális vizsgálatával („puszta szemmel") néha nehéz eldönteni, hogy az adott hálózat melyik osztályba tartozik.
- Nem mutatja a korreláció nagyságrendjét, ezért ez alapján nehéz összehasonlítani egymással a különböző nagyságú korrelációt mutató hálózatokat.
- Az $e_{k,k'}$ nagyjából $k_{\max}^2/2$ független változót tartalmaz, s ez bizonyos esetekben túl sok információ ahhoz, hogy analitikus számításokban és szimulációkban modellezhető legyen.

Szerencsére sokkal tömörebb módszerek is vannak a fokszám-korreláció mérésére, amelyek közül a legnépszerűbbet a következő alfejezetben mutatjuk be.

A szomszédok átlagos fokszáma

Egy természetes gondolat a fokszám-korreláció leírására az, hogy a problémát az élek helyett a csúcsok felől is megközelíthetjük. Ebben az esetben az a kérdés, vajon egy véletlenszerűen választott csomópont szomszédainak (átlagos) fokszáma mennyire hasonlít (korrelál) magának a csomópontnak a fokszámához. Ennek vizsgálatára először vezessük be egy adott i csúcs közvetlen szomszédainak átlagos fokszámát a

$$k_{\mathrm{nn}}^{(i)} = \frac{1}{k_i} \sum_{j \in \mathrm{nn}(i)} k_j \qquad 41.$$

alakban,[25] ahol a jobb oldalon az összegzés az i szomszédain fut végig (az „nn" index az angol nearest neighbour kifejezésre utal). Ha ezt meghatároztuk az összes csomópontra, akkor utána a csomópontokat összegyűjthetjük fokszám szerint, és megnézhetjük, hogy egy adott k fokszám esetén az ilyen fokszámmal rendelkező csúcsokra a most bevezetett $k_{\mathrm{nn}}^{(i)}$átlagosan mekkora. Ezzel megkapjuk a k fokszámú csúcsok közvetlen szomszédainak átlagos fokszámát:

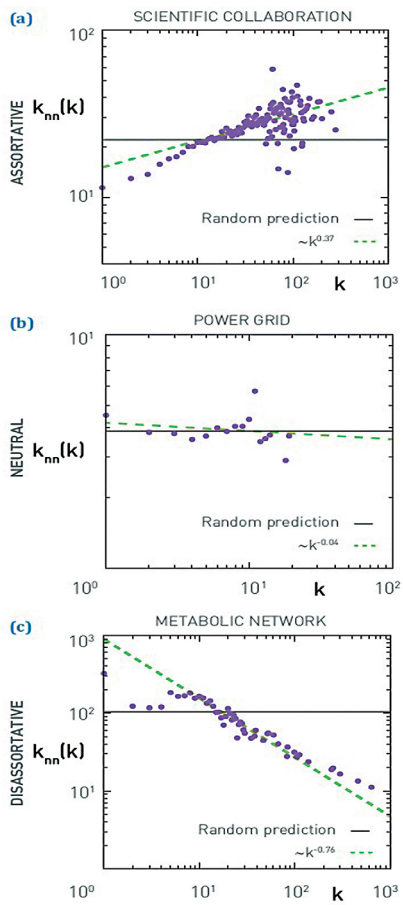
$$k_{\mathrm{nn}}(k) = \frac{1}{N_k} \sum_{i:k_i=k} k_{\mathrm{nn}}^{(i)} \qquad 42.$$

ahol N_k a k fokszámú csúcsok számát jelöli a hálózatban.

[25] PASTOR-SATORRAS–VÁZQUEZ–VESPIGNANI 2001; VÁZQUEZ–PASTOR-SATORRAS–VESPIGNANI 2002.

Nézzük meg, kvalitatíve miként viselkedik a $k_{nn}(k)$ a fokszám függvényében:

- Asszortatív hálózat esetén, mivel a kis fokszámú csúcsok többnyire kis fokszámú csúcsokkal szomszédosak, a nagy fokszámúak meg nagyokkal, a $k_{nn}(k)$ *emelkedő* tendenciát mutat a fokszám függvényében.
- Neutrális hálózat esetén a szomszédok átlagos fokszáma nagyjából független az adott csúcs fokszámától, ezért $k_{nn}(k)$ várhatóan konstans, azaz sem nem emelkedik, sem nem csökken, ha változtatjuk k-t.
- Diszasszortatív hálózat esetén, annak révén, hogy a kis fokszámú csúcsok hubokhoz, a nagy fokszámúak pedig kicsikhez szeretnek kötődni, a $k_{nn}(k)$ a fokszám függvényében *ereszkedő* tendenciát mutat.



34. ábra

A szomszédok átlagos fokszáma a k függvényében

Megjegyzés: A fokszám-korreláció mérésének egyik népszerű eszköze a $k_{nn}(k)$ görbe, amely emelkedő asszortatív hálózatok esetén, mint az a) panelen, nagyjából konstans, mint a b) panelen, és ereszkedő diszasszortatív hálózatokban, mint a c) panel esetén.

Forrás: BARABÁSI 2017

A $k_{\mathrm{nn}}(k)$ grafikonját láthatjuk három valós hálózat esetén a 34. ábrán. A tudományos társszerzőségi hálózat [34. ábra a) képe] asszortatív, és ezzel összhangban $k_{\mathrm{nn}}(k)$ emelkedő tendenciát mutat. Lévén, hogy ez egy skálafüggetlen hálózat, a fokszámok nagyon széles skálán fordulnak elő, ezért a grafikont log-log skálán tüntetjük fel. Látható, hogy az adatpontok nagyjából egy egyenest követnek az ábrán, ami természetesen egy hatványszerű függésre utal, ami alapján bevezethetjük a fokszám-korreláció erősségét jellemző μ exponenst a

$$k_{\mathrm{nn}}(k) \simeq \mathrm{a} \cdot k^{\mu} \qquad 43.$$

egyenlet segítségével[26] (ahol a egy illesztett konstans). Az ábra alapján ennél a hálózatnál ezen exponens értéke nagyjából $\mu = 0{,}37 \pm 0{,}11$.

A 34. ábra b) képén az erősáramú villamossági hálózat esetén kapott $k_{\mathrm{nn}}(k)$ görbe látható, szintén log-log skálán. Az adatpontok nem mutatnak releváns k függést, ami alapján leszűrhetjük, hogy a hálózat neutrális. A 43. függvényalak illesztése itt $\mu = 0{,}04 \pm 0{,}05$ értéket ad.

Végül a 34. ábra c) képén a metabolikus hálózathoz tartozó $k_{\mathrm{nn}}(k)$ grafikon szerepel, amely ereszkedő, azaz diszasszortatív viselkedésre utal. Itt a μ értékére a $\mu = -0{,}76 \pm 0{,}04$ eredmény adódott.

Összefoglalva: a fokszám-korreláció tanulmányozására a legcélszerűbb eszköz a $k_{\mathrm{nn}}(k)$ grafikon kirajzolása, mert ez már sokkal tömörebb leírást ad, mint a korábban ismertetett fokszám-együttelőfordulási mátrix, viszont még mindig elég részletes információkkal szolgál. Ha azonban még ennél is tömörebb mutatót szeretnénk megadni, akkor használhatjuk az imént bevezetett μ exponenst is, hiszen ha μ értéke szignifikánsan nagyobb, mint 0, akkor a hálózat asszortatív, ha μ nullához közeli értéket vesz fel, akkor a hálózat neutrális, és ha μ szignifikánsan kisebb, mint 0, akkor a hálózat diszasszortatív.

Strukturális diszasszortativitás

A skálafüggetlen hálózatok fokszám-korrelációjának vizsgálatakor fontos figyelembe venni, hogy bizonyos esetekben a hálózat mutathat ereszkedő $k_{\mathrm{nn}}(k)$ görbét (ami diszasszortatív jellegre utalna) úgy, hogy közben a különböző fokszámok közti kapcsolatok gyakorisága mégis megegyezik azzal, amit egy ugyanolyan fokszámeloszlású, de véletlenszerűen összehuzalozott hálózatban várnánk. Ennek a látszólagos ellentmondásnak a feloldása a következő: a skálafüggetlen hálózatokban extrém nagy fokszámú csúcsok is könnyűszerrel előfordulhatnak, és bizony, ha kiszámítjuk két ilyen hub között a véletlenszerű kapcsolódás esetén várt élek számát, akkor néha 1-nél jóval magasabb értékek is kijöhetnek. Ez azt jelenti, hogy ezekben a rendszerekben egy véletlenszerű (de a fokszámokat megtartó) huzalozás esetén azt várnánk, hogy az adott hubok *több éllel is kapcsolódnak egymáshoz.* Azonban ha nem multigráfot használunk az adott komplex rendszer reprezentálására, hanem egyszerű gráfot, akkor természetesen maximum 1 kapcsolatot engedünk meg, és emiatt a nagy fokszámú csúcsok között kevesebb él lesz, mint amit a neutrális, véletlenszerű kapcsolódás alapján várnánk, ami megjelenik a $k_{\mathrm{nn}}(k)$ ereszkedő viselkedésében is.

[26] Pastor-Satorras–Vázquez–Vespignani 2001.

Ezt a jelenséget hívjuk *strukturális diszasszortativitásnak,* ahol a strukturális arra utal, hogy a hálózat látszólagos diszasszortativitása szimplán a fokszámeloszlásának (és a tiltott többszörös kapcsolatoknak) a következménye. Joggal vetődik fel a kérdés, hogy miként lehet eldönteni egy ereszkedő $k_{\mathrm{nn}}(k)$ esetén, hogy a hálózat csak strukturálisan diszasszortatív, vagy valóban „nagyobb a taszítás" a nagy fokszámú csúcsok között fokszámfüggetlen kapcsolódáshoz képest? A legkézenfekvőbb ilyenkor a $k_{\mathrm{nn}}(k)$ görbét lemérni a hálózat véletlenszerűen áthuzalozott (más néven randomizált) másolataiban is. Itt most két randomizálást említünk meg:

- *Fokszám-megtartó randomizálás egyszerű gráfokkal (R-S):* a csúcsok fokszámát megőrizve véletlenszerűen huzalozzuk át az éleket úgy,[27] hogy a hálózat minden lépésben egy egyszerű gráfnak felel meg, azaz a többszörös kapcsolatok tiltottak. [Az (R-S) rövidítés az angol *random simple graph*-ra utal.]
- *Fokszám-megtartó randomizálás multigráfokkal (R-M)*: a csúcsok fokszámát megőrizve véletlenszerűen huzalozzuk át az éleket úgy, hogy megengedjük a többszörös kapcsolatok létrejöttét, azaz fennáll a lehetősége annak, hogy egy multigráf alakuljon ki. [Az (R-M) rövidítés az angol *random multi graph*-ra utal.]

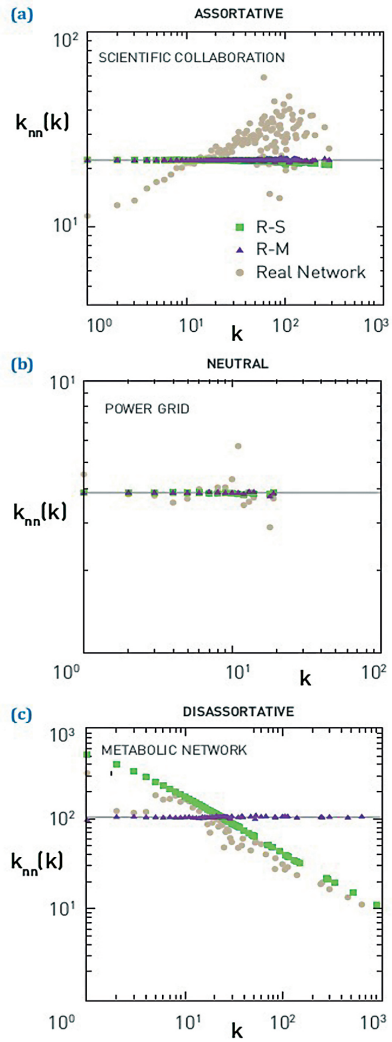
A kettő közül az (R-M) randomizáció garantáltan neutrális hálózatot hoz létre, hiszen ebben az esetben az sem probléma, ha a skálafüggetlen struktúra miatt akár több élt is be kell húzni két nagy fokszámú csúcs között. Az (R-S) randomizáció viszont csak akkor tud neutrális hálózatot létrehozni, ha nem lép fel strukturális diszasszortativitás.

Ebből azt szűrhetjük le, hogy az eredeti hálózatban mért $k_{\mathrm{nn}}(k)$ görbét elsősorban az (R-S) randomizációval kapott hálózat $k_{\mathrm{nn}}(k)$ grafikonjával kell összehasonlítanunk:

- Ha a két görbe nagyjából egybeesik, akkor a hálózat ténylegesen nem diszasszortatív, csak strukturálisan.
- Ha az (R-S) randomizáció esetén $k_{\mathrm{nn}}(k)$ már konstans, vagy sokkal kevésbé ereszkedő, mint az eredeti hálózatban, akkor a hálózat ténylegesen diszasszortatív jellemzőket mutat.

Ezt illusztrálja a 35. ábra, amelyben a 34. ábrán mutatott eredmények mellett kirajzoltuk az (R-S), illetve az (R-M) randomizációkkal kapott hálózatokban mérhető $k_{\mathrm{nn}}(k)$ görbéket is. Látható, hogy az (R-M) randomizáció minden esetben neutrális hálózatot eredményezett, ahogy ez várható. Az (R-S) randomizáció azonban neutrális hálózatot hozott létre a tudományos társszerzőségi hálózat [35. ábra a) képe] és az erősáramú villamossági hálózat [35. ábra b) képe] esetén, viszont diszasszortatív hálózatot a metabolikus hálózat esetén [35. ábra c) képe]. Ebből az következik, hogy ez a hálózat ténylegesen nem diszasszortatív, pusztán strukturális diszasszortativitás lép fel benne.

[27] A fokszámot megőrző véletlenszerű áthuzalozás módszereit bővebben a *Konfigurációs modell* című alfejezetben ismertetjük.



35. ábra

Az eredeti és a randomizált hálózatok $k_{nn}(k)$ görbéje

Megjegyzés: A strukturális diszasszortativitás vizsgálatához érdemes az (R-S) (zöld szimbólumok) és az (R-M) (lila szimbólumok) randomizáció során kapott $k_{nn}(k)$ görbéket is kirajzolni.

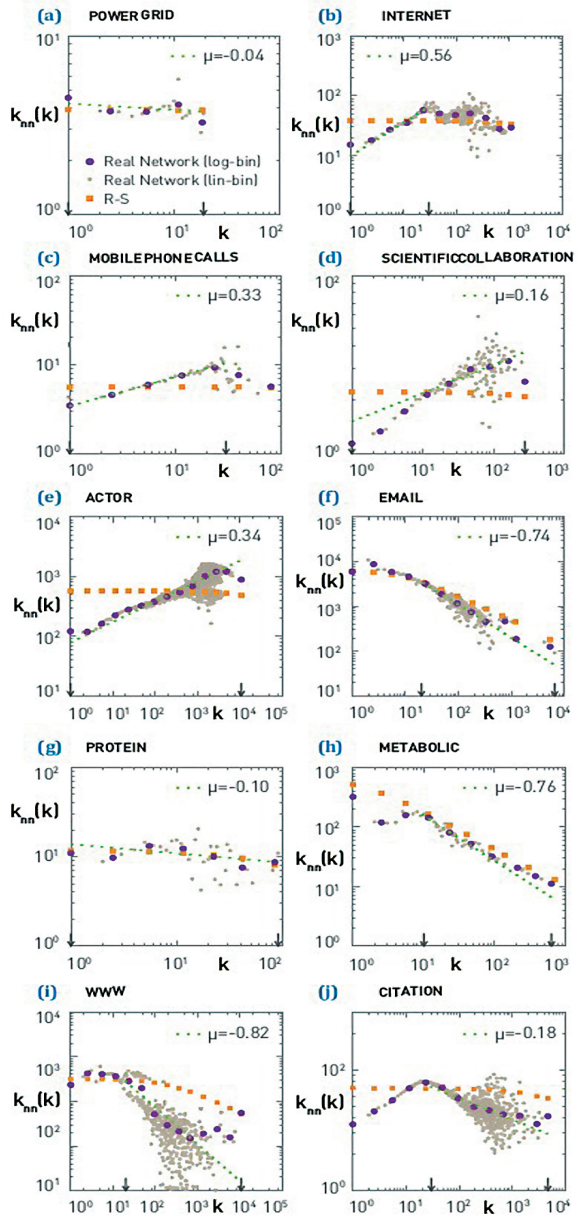
Forrás: BARABÁSI 2017

Fokszám-korreláció valós hálózatokban

Végezetül vizsgáljuk meg, hogy a mintaként használt hálózatok miként viselkednek a fokszám-korreláció szempontjából.

A 36. ábra alapján a következőket szűrhetjük le:

- Erősáramú hálózat: Az erősáramú elektromos hálózatban a $k_{\mathrm{nn}}(k)$ függvény lapos, és megkülönböztethetetlen a véletlenszerűsített változattól, nincsen tehát fokszám-korreláció [36. ábra a) kép]. Az erősáramú hálózat fokszám-korrelációja ilyenformán neutrális.
- Internet: Kis fokszámokon ($k \leq 30$) a $k_{\mathrm{nn}}(k)$ a tisztán asszortatív hálózatra jellemző módon növekszik, majd a magas fokszámokon „ellaposodik" [36. ábra b) kép]. Az internet térképének véletlenszerűsített változatában eltűnik a fokszám-korreláció. Az internet tehát asszortatív, de nagy k értékeken a strukturális levágás megszünteti ezt az asszortatív hatást.
- Szociális hálózatok: A három, szociális kapcsolatokat magában foglaló hálózat, a mobilhívások, a tudományos együttműködések és a színészek hálózata esetén $k_{\mathrm{nn}}(k)$ növekvő függvény, s ez asszortativitást jelez [36. ábra c)–e) kép]. Ennélfogva ezekben a hálózatokban a hubok főleg hubokhoz kapcsolódnak, az alacsony fokszámú csomópontok pedig főleg alacsony fokszámú csomópontokhoz. Az a tény, hogy a megfigyelt $k_{\mathrm{nn}}(k)$ eltér az (R–S) randomizáció során kapott görbétől, azt jelzi, hogy a szociális hálózatok asszortativitása nem a skálafüggetlenségük következménye.
- E-mail-hálózat: Bár az e-mail-hálózatot gyakran tekintjük szociális hálózatnak, növekvő k értékeken a $k_{\mathrm{nn}}(k)$ csökken, a hálózat tehát tisztán diszasszortatív [36. ábra f) kép]. Az (R–S) randomizáció során kapott görbe szintén csökken, ezért a látható diszasszortativitás strukturális természetű, és a hálózat skálafüggetlenségének következménye.
- Biológiai hálózatok: A fehérje-kölcsönhatások hálózatában és az anyagcsere-hálózatban μ negatív, és ez arra utal, hogy ezek a hálózatok diszasszortatívok. Az (R-S) randomizáció során kapott görbe skálázása azonban megkülönböztethetetlen az eredeti $k_{\mathrm{nn}}(k)$-től, tehát strukturális diszasszortativitást látunk, s annak a hálózat skálafüggetlensége az oka [36. ábra g)–h) kép].
- Web: Az ereszkedő $k_{\mathrm{nn}}(k)$ diszasszortatív korrelációt jelent [36. ábra i) kép]. Az (R-S) randomizáció görbéje szintén csökken, de nem olyan gyorsan, mint $k_{\mathrm{nn}}(k)$. A web diszasszortatív természetét tehát csak részben magyarázza skálafüggetlen természete.
- Idézettségi hálózat: A hálózat viselkedése elgondolkodtató: a $k \leq 20$ tartományban a $k_{\mathrm{nn}}(k)$ fokszám-korrelációs függvény tisztán asszortatívként viselkedik, de a $k > 20$ tartományban a skálázás már diszasszortatív [36. ábra j) kép]. Ez a kevert viselkedés olyan hálózatokban jelenhet meg, amelyek nagyon erős asszortativitást mutatnak. Ez arra utal, hogy az idézettségi hálózat erősen asszortatív, de a skálafüggetlensége strukturális diszasszortativitást okoz, és a nagy fokszámok tartományában megváltoztatja a $k_{\mathrm{nn}}(k)$ meredekségét.



36. ábra

Fokszám-korreláció valódi hálózatokban

Megjegyzés: A tíz referenciahálózat $k_{nn}(k)$ fokszám-korrelációs függvénye. A szürke jelek a $k_{nn}(k)$ függvényt mutatják lineáris dobozolással; a lila körök ugyanazokat az adatokat mutatják logaritmikus dobozolással. A zöld pontozott vonal a 43. függvényre való legjobb illesztésnek felel meg a lent nyilakkal jelölt intervallumban. A narancsszínű négyzetek az (R-S) randomizáció esetén mérhető $k_{nn}(k)$ görbét mutatják.

Forrás: BARABÁSI 2017

Az eredményeket összefoglalva elmondhatjuk, hogy a tíz referenciahálózatból az erősáramú villamossági hálózat az egyetlen igazi neutrális hálózat, a többi rendszerben van valamilyen fokszám-korreláció. Egy másik érdekes tapasztalat, hogy a mintaként használt hálózatokból a diszasszortatív rendszerek majdnem mind a fokszámeloszlás miatti strukturális okból diszasszortatívak. Ez alól a web volt a kivétel, ahol a fokszámeloszlás csak részben magyarázza a diszasszortativitást. Az asszortatív hálózatok esetén azonban a fokszámeloszlás nem magyarázza meg a fokszám-korrelációt. A legtöbb szociális hálózat (mobilhívások, tudományos együttműködések, színészek hálózata) idesorolható, akárcsak az internet és az idézettségi hálózat.

Bizonyos esetekben vannak természetes mechanizmusok, amelyek joggal vetődhetnek fel a megfigyelt asszortativitás eredetének magyarázására. Például az egyének közösségeket, csoportokat szoktak alkotni (ilyen például egy család, egy baráti kör vagy egy munkahelyi közösség), és ez asszortatív korrelációkhoz vezethet. Emellett maga a társadalom is a szakértői bizottságoktól kezdve a tévéműsorokig rengeteg módszerrel hozza össze akaratlanul is a hubokat, növelve ezzel a szociális és szakmai hálózatok asszortativitását. Továbbá a szociológusok által feltárt homofília is azt mutatja, hogy az egyének hasonló hátterű és jellemű egyénekkel barátkoznak és ismerkednek szívesen, aminek egy áttételes következménye lehet, hogy a rendkívül sok kapcsolattal rendelkező személyek egymással is össze vannak kapcsolva.

Véletlengráf-modellek

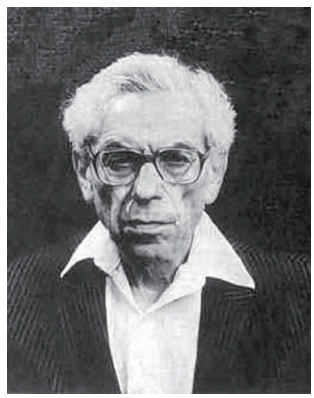
Kiemelt szerepe van a hálózatkutatás területén a véletlengráf-modelleknek. Ahogy a nevük is mutatja, ezek olyan matematikai modellrendszerek, amelyek segítségével véletlen huzalozással rendelkező gráfokat lehet létrehozni különböző szabályok szerint. Ezek a modellek egyrészt segíthetnek megérteni, hogy a társadalom és a természet komplex rendszereit leíró hálózatok bizonyos tulajdonságai milyen mechanizmusok révén állhatnak elő, valamint mindig fontos összehasonlítási alapot kínálnak egy-egy valós hálózat tanulmányozásakor. Ezenfelül a gyakorlatban a különféle hálózatokon futtatható algoritmusok (mint például keresési algoritmusok vagy a klaszterezési eljárások) tesztelésénél is nagyon hasznosak lehetnek, hiszen segítségükkel tetszőleges számban és méretben lehet teszthálózatokat előállítani, amelyek ugyan véletlenszerűek, de bizonyos tulajdonságaik felett kontrollal bírunk. Ebben a fejezetben a négy talán legalapvetőbbnek tekinthető véletlengráf-modellt ismertetjük vázlatosan.

Az Erdős Pál és Rényi Alfréd által 1959-ben bevezetett modell egy nagyon fontos mérföldkő a diszkrét matematikában és a hálózatkutatásban egyaránt.[28] Mint korábban említettük, ezt a modellt szokás *klasszikus véletlen gráfnak* is hívni, és mind a mai napig az egyik legfontosabb referenciapont bármilyen valós hálózat elemzése során.

A modell egyszerűsége miatt a legtöbb hálózatjellemző könnyen kiszámítható; ezt kihasználva a korábbi fejezetekben lépten-nyomon már használtuk is ezt a véletlen gráfot a bevezetett mennyiségek szemléltetésére. Ennek révén itt most csak röviden összefoglaljuk a modell definícióját és legfontosabb tulajdonságait, visszautalva a korábbi fejezetekben bemutatott, részletesebb magyarázatokra.

[28] ERDŐS–RÉNYI 1959; ERDŐS–RÉNYI 1960.

Erdős–Rényi-modell



37. ábra
Erdős Pál (1913–1996)
Forrás: Magyar Elektronikus Könyvtár

38. ábra
Rényi Alfréd (1921–1970)
Forrás: Magyar Elektronikus Könyvtár

Modelldefiníció és alaptulajdonságok

Az Erdős–Rényi-modellben N csomópont között minden lehetséges pontpárt egymástól függetlenül p valószínűséggel kötünk össze, létrehozva így egy véletlen gráfot. Az eredmény egy nagyon homogén hálózat lesz, amelyben minden pontnak ugyanakkora esélye van arra, hogy kapcsolatokat szerezzen magának, és ennek révén, ha bármely csúcs tetszőleges tulajdonságát vizsgáljuk (például fokszám, klaszterezettségi együttható stb.), az adott tulajdonságot jellemző számérték nagyon közel fog esni az átlagos értékhez, és nem fogunk kiugró eseteket találni. Az Erdős–Rényi-féle véletlen gráf alapvető jellemzői:

- Az élek várható száma

$$\langle L \rangle = pN(N-1)/2 \qquad 44.$$

  hiszen összesen $N(N-1)/2$ különböző pontpár létezik, és mindegyik függetlenül p valószínűséggel jön létre.
- Az átlagos fokszám

$$\langle k \rangle = (N-1)p \simeq Np \qquad 45.$$

(a részletesebb tárgyalását a *Fokszám* című alfejezetben mutattuk be).

- Az átlagos távolság (*Távolság* című alfejezet):

$$\langle d_{ij} \rangle \simeq \ln N / \ln \langle k \rangle \qquad 46.$$

- Az átlagos klaszterezettségi együttható (*Távolság* című alfejezet):

$$\langle C \rangle \simeq p \tag{47.}$$

- A fokszámeloszlás binomiális, ami általában jól közelíthető Poisson-eloszlással (*A véletlen gráf fokszámeloszlása* című alfejezet 18. és 19. egyenletek):

$$p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \simeq \frac{\langle k \rangle^k}{k!} e^{-\langle k \rangle} \tag{48.}$$

- Óriás komponens akkor van a hálózatban, ha $\langle k \rangle \geq 1$ (*Óriás összefüggő komponens* című alfejezet, 20. ábra).

Az Erdős–Rényi-gráf és a valós hálózatok

A hálózatjellemzők bemutatása során négy univerzális tulajdonságot tárgyaltunk, amelyek a valódi komplex rendszereket leíró hálózatokra többnyire teljesülnek. Nézzük meg röviden, hogy ezek közül melyeket lehet az Erdős–Rényi-gráffal reprodukálni.

- *Ritkaság: A valós hálózatok ritkák* című alfejezetben láttuk, hogy a valós hálózatok túlnyomó többsége ritka. Ezt a tulajdonságot az Erdős–Rényi-modellben egy egyszerű paraméterbeállítás segítségével tudjuk reprodukálni. Adott N méretű és $\langle k \rangle$ átlagos fokszámú valós hálózat modellezésekor a p paramétert a 38. egyenlet alapján a $p = \langle k \rangle / N$ értékre állítjuk be, és így a kapott átlagos fokszám meg fog egyezni az eredeti valós hálózat átlagos fokszámával.
- *Kis világ tulajdonság: A kis világ tulajdonság* című alfejezetben bemutattuk, hogy a valós hálózatok túlnyomó többsége kis világ tulajdonságú. Ugyanitt azt is részleteztük, hogy ez az Erdős–Rényi-gráfra is teljesül. Talán azt érdemes megjegyezni ezen a téren, hogy nagyon nagy méretű skálafüggetlen valós hálózatok esetén előfordulhat, hogy a skálafüggetlenségből következő „ultra kis világ" tulajdonság miatt a valódi hálózatban némileg kisebb átlagos távolság adódik, azonban ez az eltérés még a nagyon nagy méretű valós hálózatok esetén sem lesz számottevő.
- *Magas klaszterezettségi együttható: A valós hálózatok magasan klaszterezettek* című alfejezetben láthattuk, hogy a valós hálózatokban magas az átlagos klaszterezettségi együttható, azaz annak ellenére, hogy globálisan ritkák, lokálisan sűrűek. Ezzel szemben az Erdős–Rényi-gráf teljesen homogén, benne a globális és a lokális élsűrűség megegyezik, ennélfogva a ritka Erdős–Rényi-gráfoknak alacsony a klaszterezettségi együtthatójuk. A magas klaszterezettségi együttható tehát egy olyan univerzális tulajdonsága a valós hálózatoknak, amit az Erdős–Rényi-gráf nem képes reprodukálni.
- *Skálafüggetlenség: A skálafüggetlenség univerzális* című alfejezetben megtárgyaltak szerint a valós hálózatok túlnyomó többsége skálafüggetlen, azaz a fokszámeloszlásuk a nagy fokszámok tartományán hatványszerűen cseng le, $p(k) \sim k^{-\gamma}$, ami rendkívüli inhomogenitást okoz a csúcsok fokszámai között, hubok megjelenéséhez vezet, és ezek nagyon jelentősen befolyásolják például a hálózat robusztusságát vagy a terjedési jelenségek lefolyását. Mivel az Erdős–Rényi-gráf esetén binomi-

ális (Poisson) eloszlást követ a $p(k)$, értelemszerűen ezt a tulajdonságot sem képes reprodukálni, és a valós hálózatokkal szemben a létrejövő véletlen gráf homogén.

Összefoglalva azt láthatjuk, hogy az Erdős–Rényi-modell sok tekintetben jelentősen eltér a valós komplex rendszereket leíró hálózatok struktúrájától. Ennek ellenére nagyon fontos pontja a hálózatelméletnek, hiszen olyan viszonyítási alapot kínál, amelyre támaszkodva jobban megérthetjük a valós hálózatok jellemzőit.

Watts–Strogatz-modell

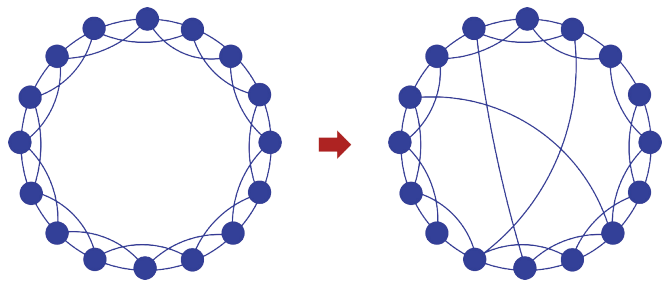
39. ábra
Duncan J. Watts
Forrás: Columbia News

40. ábra
Steven Strogatz
Forrás: Speakerpedia

Az első híres alternatívája a klasszikus véletlen gráfnak a Duncan J. Watts és Steven Strogatz által 1998-ban javasolt modell volt.[29] A valós hálózatokban tapasztalt magas átlagos klaszterezettségi együttható által motiválva a fő cél egy olyan véletlengráfmodell bevezetése volt, amely egyszerre képes kis világ tulajdonságú és magasan klaszterezett véletlen hálózatot előállítani (úgy, hogy közben a kapott hálózat ritka marad).

A modell alapötlete az, hogy induljunk ki egy olyan szabályos gráfból, amelyben biztosított a magas klaszterezettség, majd ebben az élek kis hányadának véletlenszerű átkötésével csökkentsük le az átlagos távolságot, hogy a kis világ tulajdonság is teljesüljön.

[29] Watts–Strogatz 1998.



41. ábra

A Watts–Strogatz-modell

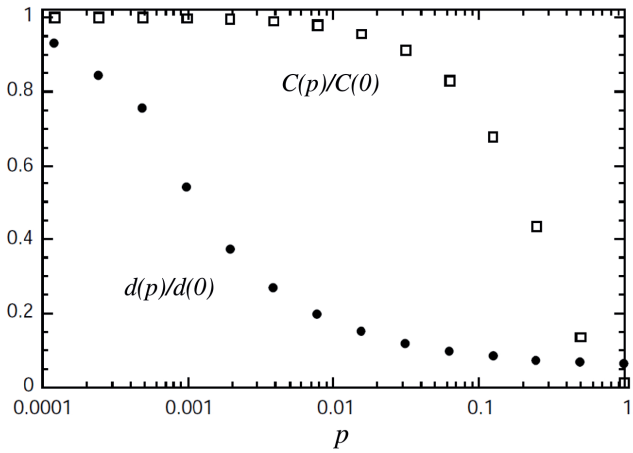
Megjegyzés: Ebben a modellben egy szabályos kör mentén elhelyezett pontokból álló gráfból indulunk ki, ahol minden csomópontnak közvetlen kapcsolata van a kör mentén hozzá legközelebb eső q ponthoz mindkét irányban. Az ábra bal oldalán a $q = 2$ esetet láthatjuk. A szabályos, induló gráfban minden élt egy rögzített p valószínűséggel véletlenszerűen átkötünk. Amíg p nem túl nagy, ez az éleknek csak kis hányadát érinti, ahogy azt a jobb oldali gráfon láthatjuk.

Forrás: a szerző szerkesztése

A javasolt alaphálózat egy N csúcsból álló kör volt, amelyben a kör mentén minden csomópont közvetlen kapcsolatot létesített nemcsak az első, hanem egészen a q-adik szomszédjáig, ahogy azt a 41. ábra bal oldalán láthatjuk. (Természetesen hálózati értelemben ezek így mind közvetlen szomszéddá váltak, a 2., 3., …, q. szomszéd elnevezés itt a kör mentén értendő.) Ha $q > 1$, akkor ezzel automatikusan sok háromszöget hozunk létre a hálózatban, amitől a klaszterezettségi együttható magas. [Belátható, hogy a klaszterezettségi együttható értéke $C = (3q - 3)/(4q - 2)$.] Azonban ebben a szabályos hálózatban az átlagos legrövidebb úthossz a csúcsok számával arányos $\langle d \rangle \sim N$, hasonlóan egy szimpla körhöz vagy lánchoz.

Ahhoz, hogy lecsökkentsük az átlagos távolságot, és kis világgá tegyük ezt a rendszert, elég, ha a kapcsolatok kis hányadát véletlenszerűen átkötjük, ami technikailag azt jelenti, hogy végigmegyünk az élek listáján, és egységesen p valószínűséggel az egyik végpontjukat átkötjük egy véletlenszerűen választott csomópontra (ahol természetesen p egy viszonylag kis érték). Ezt illusztrálja a 41. ábrán a jobb oldalon látható hálózat, amelyben 3 élt kötöttünk át.

A 42. ábrán láthatjuk, hogy miként alakul a kapott véletlen gráfban az átlagos klaszterezettség és az átlagos távolság a p átkötési valószínűség függvényében. Mindkét mennyiség esetén a mérhető átlag le lett osztva az átkötés nélküli kiinduló állapotban mért értékkel annak érdekében, hogy a két mennyiség kényelmesen ábrázolható legyen egy görbén. Látható, hogy az átlagos távolság görbéje viszonylag hamar gyors ereszkedésnek indul, és a hálózat eléri a kis világ állapotot. Ezzel szemben a klaszterezettségi együttható grafikonja sokáig szinte alig változik, és csak sokkal később kezd el csökkenni. Ennek köszönhetően a p értéknek van egy széles tartománya, amelyen belül a kapott hálózat egyszerre magasan klaszterezett és kis világ tulajdonságú.



42. ábra

Az átlagos távolság és az átlagos klaszterezettségi együttható viselkedése a Watts–Strogatz-modellben

Megjegyzés: A grafikon vízszintes tengelyén a p átkötési valószínűség van feltüntetve. A fekete körök a mért átlagos $d(p)$ távolságot mutatják, a kezdeti (átkötés nélküli) állapotban mért átlagos távolság értékével leosztva. A fehér négyzetek a mért átlagos $C(p)$ klaszterezettségi együtthatónak felelnek meg, amely szintén le lett osztva az átkötés nélküli $p = 0$ állapotban mérhető $C(0)$ értékkel.

Forrás: Watts–Strogatz 1998

Intuitíven a modell működésének az a magyarázata, hogy már viszonylag kevés számú véletlen átkötés is eredményezhet olyan hidakat a hálózatban, amelyek korábban egymástól nagyon messze lévő pontokat kötnek össze közvetlenül, drasztikusan csökkentve ezzel az átlagos távolságot.

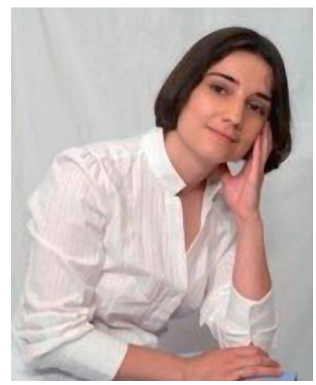
Ezzel párhuzamosan, ha a klaszterezettségi együtthatót vizsgáljuk, akkor kevés számú átkötés nem elég egy drasztikus változáshoz, hiszen a kezdeti állapotban található sok háromszög megszüntetéséhez sok átkötés is kell. Emiatt a klaszterezettségi együttható magas értéke bőven azután is fent tud maradni, hogy a kis világ effektus már létrejött a hálózatban.

Bár ez a modell kétségtelenül egy fontos mérföldkő a hálózatkutatás történetében, manapság nem szokás valós hálózatok modellezésére használni. Ennek egyszerű oka az, hogy a létrejövő véletlen gráf nem skálafüggetlen. A kezdeti állapotban minden csomópontnak azonos a fokszáma, és a véletlen átkötések ehhez képest csak minimális eltéréseket okoznak.

Barabási–Albert-modell

Az első véletlengráfmodell, amely képes egy egyszerű mechanizmus segítségével skálafüggetlen hálózatot generálni, a Barabási Albert-László és Albert Réka által bevezetett modell, amely korunk egyik leghíresebb hálózatmodellje.[30] Ez a megközelítés több aspektusában is drasztikusan eltér a korábban tárgyalt modellektől. Ezek közül az első az, hogy a valós rendszerek megfigyelésére alapozva nem statikus rendszerérettel dolgozik, hanem egy időben *növekvő* hálózatot feltételez, amelyhez minden időlépésben új csomópontok csatlakoznak, és hozzákapcsolódnak a hálózat már létező részéhez.

43. ábra
Barabási Albert-László
Forrás: CEU

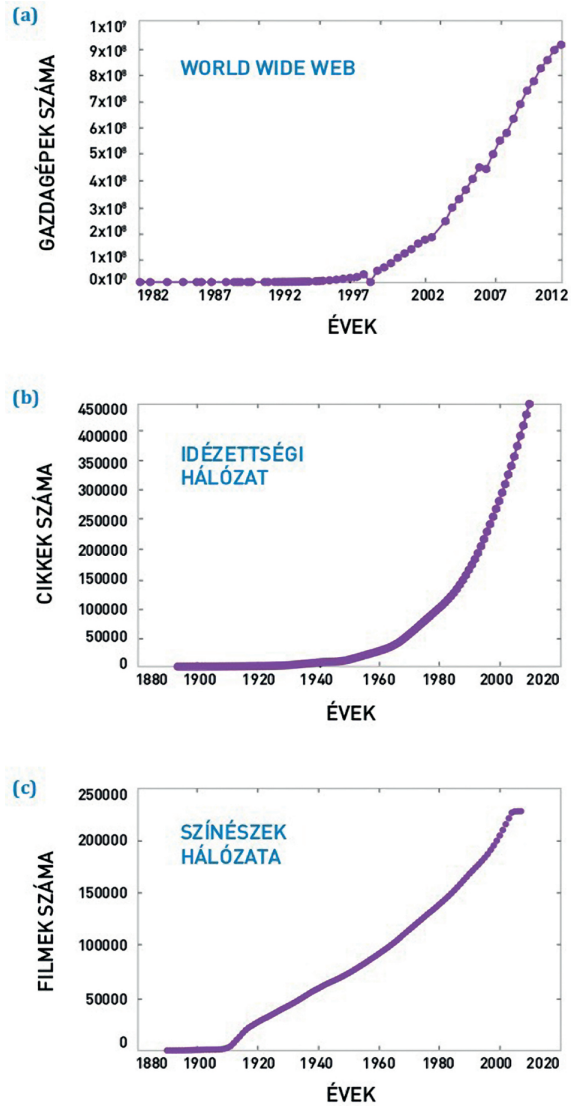


44. ábra
Albert Réka
Forrás: ScienceWatch

A 45. ábrán néhány példát láthatunk időben növekvő komplex rendszerekre, amelyeket széles körben szokás hálózatként reprezentálni, és azon hálózatok listája, amelyekről biztosan tudjuk, hogy méretük idővel növekszik, tovább bővíthető:

- 1991-ben a WWW egyetlen csomópontból állt, azt Tim Berners-Lee hozta létre. Ma már több mint egybillió (10^{12}) dokumentum alkotja. Ezt a hihetetlenül nagy számot úgy értük el, hogy milliónyi egyén és intézmény ad hozzá újabb és újabb dokumentumokat a WWW-hez [45. ábra a) kép].
- Az együttműködési és az idézettségi hálózat folyamatosan növekszik az új kutatási tanulmányok publikálásával [45. ábra b) kép].
- A színészek kapcsolati hálózata is folyamatosan nő az új filmek megjelenésével [45. ábra c) kép].
- A fehérje-kölcsönhatási hálózat állandónak tűnhet, hiszen génjeinket (és emiatt fehérjéinket) a szüleinktől örököljük. Valójában azonban ez a hálózat sem állandó, mert napjainkra – az eltelt több mint négymilliárd év alatt – az emberi sejtek génállománya néhányról húszezerre nőtt.

[30] BARABÁSI–ALBERT 1999.



45. ábra

Növekvő valós hálózatok

Megjegyzés: Sok hálózat időben nem állandó, hanem új csomópontok csatlakozásával növekedik. a) A webhelyeket tároló gépek számának változása a web gyors növekedését mutatja. b) A *Physical Review* folyóiratban megjelent tudományos cikkek száma az alapítás óta. Az egyre több cikk által bővül a tudományos együttműködési hálózat és az idézettségi hálózat. c) A filmek számának növekedése bővíti a színészhálózatot (az IMDB.com alapján).

Forrás: Barabási 2017

A másik nagyon fontos újítása a Barabási–Albert-modellnek az, hogy az új csomópontok a csatlakozáskor ugyan véletlenszerűen választanak, de *nem egyenletes* valószínűséggel a már létező pontok közül. Ez a fajta gondolat egyáltalán nincs jelen sem az Erdős–Rényi-modellben (ahol minden csúcspár között egyforma valószínűséggel jön létre él), sem a Watts–Strogatz-modellben (ahol az átkötődő élek egyenletes valószínűség szerint választanak maguknak új csomópontot). Nézzük meg, hogy mit tudunk mondani ebből a szempontból a valódi komplex rendszereket reprezentáló hálózatok esetén néhány példán keresztül:

- Az interneten elérhető egybillió dokumentumnak csak kis részét ismerjük. Az általunk ismert csomópontok nem teljesen véletlenszerűek: mindenki hallott már a Google-ről és a Facebookról, de csak ritkán találkozunk a webet alkotó milliárdnyi, kevésbé ismert csomóponttal. Mivel ismereteink a népszerűbb oldalakra korlátozódnak, ezért inkább kapcsolódunk egy magas fokszámú csomóponthoz, mint egy olyanhoz, amely csak kevés számú más csomóponthoz kapcsolódik.
- Egyetlen kutató sem képes az évente megjelent több milliónyi tudományos publikációt elolvasni. De minél idézettebb egy tanulmány, annál biztosabb, hogy hallottunk róla, vagy talán már olvastuk is. S mihelyt idézzük is azt, amit olvastunk, idézetünk úgy tolja azt a többet idézett publikációk irányába, közelíti az idézettségi hálózat magas fokszámú csomópontjai felé.
- Minél több filmben játszott egy színész, annál jobban ismeri a képességeit a szerepeket osztó rendező. Minél nagyobb egy színész fokszáma, annál valószínűbb, hogy újabb szerepet fog kapni.

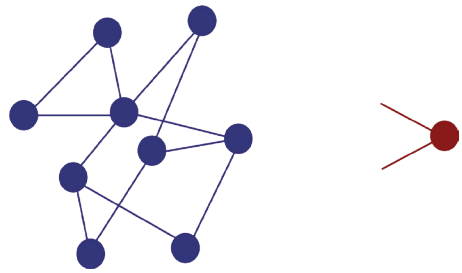
A fentiek alapján sejthető, hogy több valós hálózat esetén az újonnan megjelenő pontok sokkal nagyobb eséllyel kapcsolódnak a már létező pontok között a nagy fokszámú csomópontokhoz, mint a kis fokszámúakhoz. Ezt a mechanizmust hívjuk *preferenciális kapcsolódásnak,* és a növekedés mellett ez képezi a Barabási–Albert-modell másik sarokkövét.

Modelldefiníció és a fokszámeloszlás

Magát a Barabási–Albert-modellt a következő lépéseken keresztül definiálhatjuk:

- A hálózat kezdeti állapota egy m_0 számú csomópontból álló, összefüggő hálózat. (Ez lehet egy véletlen gráf, egy lánc, egy fa stb., ennek a belső szerkezete nincs lényeges hatással a hálózat hosszú távú fejlődésére).
- Minden időlépésben 1 új csomópontot adunk hozzá a hálózathoz, amely m új kapcsolatot létesít a már létező csúcsok felé, ahol természetesen $m \leq m_0$ (46. ábra).
- Az új kapcsolatok „szabad” végének bekötésekor a már létező csomópontok közül a f*okszámukkal arányos valószínűséggel* választunk, ez a preferenciális kapcsolódási szabály. Ez alapján annak valószínűsége, hogy az i csomóponthoz fog kapcsolódni egy új él:

$$P\,(\text{az } i\text{-t választjuk}) = \frac{k_i}{\sum_j k_j} \qquad 49.$$



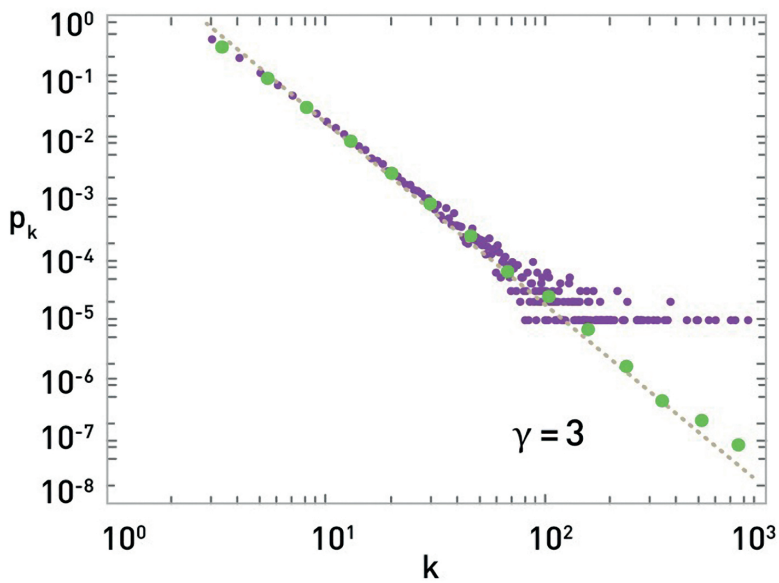
46. ábra

A Barabási–Albert-modell szemléltetése

Megjegyzés: Minden időlépésben egy új csomópont érkezik a hálózatba, amely m új éllel kapcsolódik hozzá a hálózat már létező részéhez. Az új csomópontot és az új éleket pirossal jelöltük; a jelen esetben $m = 2$.

Forrás: a szerző szerkesztése

Ha ezen szabályok alapján generálunk hálózatot, akkor kellően nagy rendszerméret fölött a fokszámeloszlás skálafüggetlenné válik, ahogy azt a 47. ábra mutatja.



47. ábra

Egy Barabási–Albert-modell alapján szimulált hálózat fokszámeloszlása

Megjegyzés: A csomópontok száma $N = 100000$, az m paraméter értéke (az új csúcsokból induló új élek száma) $m = 3$. A lila az eredeti adatpontokat mutatja, a zöld a logaritmikus dobozolással kapott eredményt, az egyenes vonal pedig egy $\gamma = -3$ exponenssel csökkenő hatványfüggvénynek felel meg.

Forrás: Barabási 2017

A fokszámeloszlás hatványszerű lecsengését egy közelítő analitikus levezetésből is megkaphatjuk, amelynek előnye, hogy a γ exponens értékét is megadja. Ehhez az egyszerűség kedvéért kezeljük az i csúcs $k_i(t)$ fokszámát egy folytonos változóként, amelynek várható kis megváltozása a t és $t+1$ időlépések között

$$\Delta k_i(t) = \frac{k_i(t)}{\sum_j k_j(t)} \cdot m \qquad 50.$$

ahol behelyettesítettük a 49. valószínűséget, és kihasználtuk, hogy összesen m új él csatlakozik be a hálózatba egy időlépés alatt. Ha egy mennyiség változásának nagyságát elosztjuk az időlépés hosszával, akkor megkapjuk az időlépés által elválasztott két időpontban felvett értékekre illesztett egyenesnek a meredekségét. Mi az időegységet tetszőlegesen választhatjuk, hiszen ez a modell nem „valós” időben zajlik, ezért az egyszerűség kedvéért válasszunk $\Delta t = 1$-et, amely által

$$\frac{\Delta k_i(t)}{\Delta t} = \frac{k_i(t)}{\sum_j k_j(t)} \cdot m \qquad 51.$$

Ez az egyenlet tehát megadja a $k_i(t)$ görbének a t és t+1 időpont közötti meredekségét. Az egyszerűség kedvéért tekintsünk erre úgy, mint a $k_i(t)$ időderiváltjára, ami alapján a következő egyszerű differenciálegyenletre jutunk a $k_i(t)$-re nézve:

$$\frac{\mathrm{d}k_i}{\mathrm{d}t} = \frac{k_i(t)}{\sum_j k_j(t)} \cdot m \qquad 52.$$

Tegyük fel, hogy egy nagyon nagy hálózatot növesztünk, amelyben a kezdeti m_0 számú csomópont csak egy nagyon kicsi, elhanyagolható hányadot képvisel. Ilyenkor a csomópontok számát közelíthetjük az időlépések számával (hiszen minden időlépésben 1 új csúcs jelent meg), $N \simeq t$ és hasonlóan, az élek számát közelíthetjük úgy, mint $L \simeq mt$ (hiszen minden időlépésben m új él jelent meg). Kihasználva, hogy egy hálózatban a fokszámok összege az élek számának duplája, az 52. egyenletet a következő módon alakíthatjuk át:

$$\frac{\mathrm{d}k_i}{\mathrm{d}t} = \frac{k_i(t)}{2L(t)} \cdot m = \frac{k_i(t)}{2mt} \cdot m = \frac{k_i(t)}{2t} \qquad 53.$$

Ennek megoldása

$$k_i(t) = ct^{1/2} \qquad 54.$$

ahol c egy konstans, amelynek értékét egy határfeltétel segítségével lehetne rögzíteni. Szerencsére épp akad egy ilyen időbeli határfeltétel: abban az időpontban, amikor i megjelenik a hálózatban mint új csomópont, a fokszáma nem lehet más mint $k_i = m$. Jelöljük t_i-vel az i megjelenésének időpontját; ez alapján

$$k_i(t=t_i)=m=ct_i^{1/2} \rightarrow c=\frac{m}{t_i^{1/2}} \qquad 55.$$

Ezt visszahelyettesítve

$$k_i(t)=m\left(\frac{t}{t_i}\right)^{1/2} \qquad 56.$$

A fokszámeloszlást legegyszerűbb az 56. egyenlet segítségével a $P(k)$ kumulatív eloszlásból származtatni, amely definíciószerűen adott k esetén annak valószínűsége, hogy a hálózatban egy véletlenszerűen választott csomópont fokszáma kisebb mint k, azaz

$$P(k)=P(\text{v.v.}i \text{ pont fokszáma} < k)=P(k_i<k). \qquad 57.$$

Helyettesítsük be most a k_i-re kapott 56. egyenletkifejezést:

$$P(k)=P(k_i<k)=P\left(m\left(\frac{t}{t_i}\right)^{1/2}<k\right)=P\left(\frac{t}{t_i}<\frac{k^2}{m^2}\right)=P\left(\frac{t_i}{t}>\frac{m^2}{k^2}\right) \qquad 58.$$

Az utolsó kifejezés alapján tehát a $P(k)$ kumulatív eloszlás megegyezik annak valószínűségével, hogy t_i/t nagyobb, mint m^2/k^2. Lévén, hogy $t_i \leq t$, a t_i/t mindig 0 és egy 1 közé esik. Ezenfelül, mivel az i csomópontra nézve semmilyen speciális megkötésünk nem volt, az i egyforma valószínűséggel születhetett bármelyik időlépésben 0 és t között, ezért t_i/t-re tekinthetünk úgy, mint egy a [0,1] intervallumon egyenletesen választott véletlen számra (48. ábra).

48. ábra

A t_i/t mint valószínűségi változó

Megjegyzés: Az *i* pont bármelyik lehet az *N* közül, ezért (ha *N* elég nagy, akkor) t_i/t-re tekinthetünk úgy, mint a [0,1]-en egyenletesen választott véletlen valós számra.

Forrás: xxxxxxxxxxxxxx

Azonban annak valószínűsége, hogy egy a [0,1]-en egyenletesen választott véletlen szám nagyobb, mint egy tetszőleges $x \in [0,1]$ érték, az egyszerűen $1-x$, ami alapján

$$P(k)=P\left(\frac{t_i}{t}>\frac{m^2}{k^2}\right)=1-\frac{m^2}{k^2} \qquad 59.$$

A kumulatív eloszlás a $p(k)$ fokszámeloszlás közti összefüggés folytonos változók esetén

$$P(k) = \int_0^k p(k')\mathrm{d}k', \quad p(k) = \frac{\mathrm{d}P(k)}{\mathrm{d}k} \tag{60.}$$

ezért a mi esetünkben a fokszámeloszlásra a

$$P(k) = \frac{\mathrm{d}P(K)}{\mathrm{d}k} = \frac{2m^2}{k^3} \tag{61.}$$

eredmény adódik.

Látható, hogy ebből a levezetésből egy skálafüggetlen fokszámeloszlás jött ki, amelynél a γ exponens értéke $\gamma = 3$. Noha a számolásunk tartalmazott közelítéseket, a skálafüggetlenségre és az exponens értékre kapott eredmények egzakt módon is bizonyíthatók.

További tulajdonságok

A Barabási–Albert-modell további tulajdonságait egy rövid felsorolás keretében vázoljuk, a korábban többször tárgyalt univerzális hálózatjellemzők segítségével:

- *Ritkaság: A* Barabási–Albert-modellben a hálózat átlagos fokszáma lényegében m, hiszen minden új csúcs ennyi éllel kapcsolódik hozzá a rendszerhez. Mivel ebben a rendszerben mindig $m << N$ paraméterekkel szokás dolgozni, a ritkaság automatikusan teljesül.
- *Kis világ tulajdonság:* A skálafüggetlen hálózatok tárgyalásánál, *A skálafüggetlenség jelentése és következményei* című alfejezetben említettük, hogy a $\gamma = 3$ exponenssel rendelkező hálózatokban az átlagos távolság $\langle d \rangle \sim \ln N/(\ln \ln N)$, azaz a kisvilág- és az „ultra kis világ” viselkedés határán vannak. Ez igaz a Barabási–Albert-modellre is, hiszen a γ értéke itt is $\gamma = 3$.
- *Átlagos klaszterezettségi együttható:* További analitikus számításokkal belátható, hogy az átlagos klaszterezettségi együttható $\langle C \rangle \simeq m(\ln N)^2/(8N)$, azaz vezető rendben $\langle C \rangle \sim (\ln N)^2/N$. Ez ugyan közelebb van a valós hálózatok viselkedéséhez, mint például az Erdős–Rényi-gráf (ahol $\langle C \rangle = p = \langle k \rangle/N$ azaz $\langle C \rangle \sim 1/N$), de nagy rendszerméretek esetén ez is jelentősen kisebb értéket ad, mint amit egy valós hálózatban mérnénk.

Összefoglalva: a Barabási–Albert-modell a legsikeresebb a társadalom és a természet komplex rendszereit leíró hálózatok univerzális tulajdonságainak kvalitatív reprodukálásában az eddig tárgyalt modellek közül, és ez az egyetlen, amely skálafüggetlen hálózatot generál.

A preferenciális kapcsolódás eredete

A Barabási–Albert-modell egyik alapgondolata a preferenciális kapcsolódás, amely alapján az újonnan megjelenő csomópontok a már létező pontok közül a fokszámmal arányos

valószínűséggel választanak. Ez a mechanizmus intuitív lehet egy szociális hálózatban, hiszen vannak sztárok és híresség, akikről szinte mindenki hallott már, és ez alapján nem meglepő, hogy amikor például egy új Twitter-felhasználó kiválasztja, hogy kiket fog követni, a sztárokat és híresség nagyobb eséllyel fogja bejelölni, mint egy kevés követővel rendelkező, számára ismeretlen embert. Azonban egy hasonló mechanizmus egyáltalán nem intuitív például gének vagy fehérjék hálózatában, hiszen azt senki sem gondolhatja komolyan, hogy a géneknek vagy fehérjéknek „tudomása" lenne a többiek fokszámáról, és ez alapján választanának.

A fentiek alapján érdemes egy kicsit elidőzni olyan kapcsolódási folyamatoknál, amelyek hatásukban megegyeznek a fokszám szerinti preferenciális kapcsolódással, azonban mégsem szükséges hozzájuk feltételezni azt, hogy a bekapcsolódó új csúcsoknak bármilyen információja volna a többiek fokszámáról.

- *Másoló mechanizmusok*.[31] Egy növekvő génhálózat esetén egy egyszerű plauzibilis modell az, hogy a rendszer génduplikációk folytán nő, azaz időről időre egy-egy véletlenszerűen választott gén valamilyen mutáció folytán megkettőződik. Ez annyit jelent, hogy az adott génről egy olyan módosított másolat készül, amely elég különböző ahhoz, hogy már egy másik (új) génként tekintsünk rá. Természetesen a hasonlóság még mindig relatíve nagy marad, amit hálózatosan úgy vehetünk figyelembe, hogy az új gén megörökli az eredeti gén kapcsolatait, azaz ő is ugyanazokhoz a további génekhez fog hozzákapcsolódni, mint az eredeti gén, amelyről másolódott. Ebbe a modellbe első ránézésre nem építettük bele a preferenciális kapcsolódást, mégis igaz lesz rá, hogy a nagy fokszámú géneknek sokkal nagyobb esélye lesz új kapcsolatok létesítésére. Ennek egyszerű oka az, hogy az egyetlen mód ebben a modellben a fokszám növelésére az, ha az adott csomópont szomszédja duplikálódik. Azonban a huboknak sokkal több szomszédjuk van, mint egy kis fokszámú csúcsnak, és kvantitatíven is annak esélye, hogy egy csomópont szerezzen magának egy új élt, épp a szomszédok számával (vagyis a fokszámmal) arányos. Ez az egyszerű csomópontmásoláson alapuló modell tehát egyfajta rejtett preferenciális kapcsolódást hordoz.
- *Élválasztó modellek*.[32] Növeszthetünk egy véletlen hálózatot úgy is, hogy az új csúcsok nem a már létező csúcsok közül választanak, hanem a már létező kapcsolatok közül, és egyenletes valószínűséggel. Ha megvan a kiválasztott él, akkor az új csúcsot hozzákötjük az él két végpontjához. Ez a modell sem tételez fel semmi ismeretet a fokszámokról, mégis a preferenciális kapcsolódás egyik formáját nyújtja, hiszen a nagy fokszámú csomópontokból több él indul, ezért nagyobb eséllyel szereznek további új kapcsolatokat (hiszen nagyobb a valószínűsége, hogy a kiválasztott, már létező él hozzájuk tartozik).
- *Szomszédválasztó modellek*.[33] Effektíve preferenciális kapcsolódást jelent egy olyan növekvő modell is, amelyben az új csomópont először egyenletesen véletlenszerűen választ a már létező pontok közül, majd hozzákapcsolódik ezen csomópont egy véletlenszerűen választott közvetlen szomszédjához.

[31] Kleinberg et al. 1999; Kumar et al. 2000.
[32] Dorogovtsev–Mendes 2003.
[33] Dorogovtsev–Mendes 2003.

A Barabási–Albert-modellen alapuló további modellek

A Barabási–Albert-modell a megjelenését követően hamar nagyon népszerűvé vált, és az évek során számos olyan további hálózatmodellt vezettek be, amelyek ezen a megközelítésen alapulnak, de bizonyos kisebb részletekben eltérnek az eredeti modelltől. Ezek közül itt most a fontosabbakat vázoljuk röviden.

- *Állítható exponenssel rendelkező modell. A skálafüggetlenség univerzális* című alfejezetben láthattuk, hogy a valós hálózatokban a γ exponens általában 2 és 3 közötti értéket vesz fel, és hálózatról hálózatra változik. Ezzel szemben a Barabási–Albert-modell eredeti verziója fixen $\gamma = 3$ exponensű hálózatot generál. Annak érdekében, hogy más exponensértékkel rendelkező skálafüggetlen gráfot tudjunk generálni alapvetően ugyanazon modell keretében, elég egy apró módosítás a preferenciális kapcsolódás szabályában.[34] Ha egy már létező csomópont kiválasztásának valószínűsége nem k-val, hanem $(k - a)$-val arányos (ahol a egy konstans, amelynek értéke $a \in [0,m]$), azaz

$$P(\text{az } i\text{-t választjuk}) = \frac{k_i - a}{\sum_j (k_j - a)} \qquad 62.$$

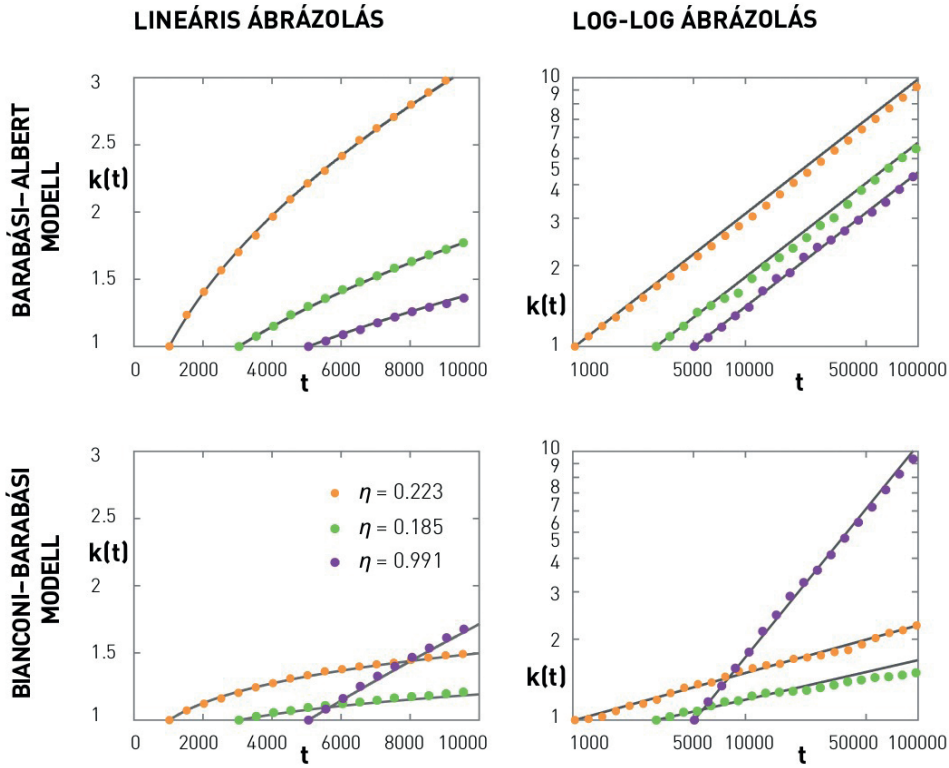
  akkor a kapott fokszámeloszlás továbbra is skálafüggetlen, azonban a γ exponens értéke $\gamma = 3 - a/m$-re módosul. Ez alapján az m és a megfelelő választásával tetszőleges γ exponenssel rendelkező skálafüggetlen hálózatot elő tudunk állítani a $2 \leq \gamma \leq 3$ tartományban.
- *Fitneszalapú, preferenciális kapcsolódású modell.* A Barabási–Albert-modell egy eddig nem részletezett tulajdonsága, hogy minél korábban jelenik meg egy csomópont a hálózatban, annál nagyobb az előnye a később érkezőkkel szemben az új kapcsolatok begyűjtésének terén, és ezáltal a legelső csúcsokból lesznek általában a legnagyobb hubok a folyamat végén. A valós hálózatokban ezzel szemben bizonyos esetekben előfordul, hogy egy-egy később érkező csomópont idővel beelőzi a nála sokkal öregebb hubokat is; gondoljunk például a Google vagy a Facebook sikerére a világhálón. Egy ilyen jelenség modellezéséhez már figyelembe kell venni a csomópontok „belső adottságait" is, amit most az egyszerűség kedvéért nevezzünk fitnesznek. A Ginestra Bianconi és Barabási Albert-László által javasolt modellben a csomópontok fitneszét egy véletlen (rögzített eloszlásból sorsolt) változó reprezentálja,[35] és a preferenciális kapcsolódási szabály annyiban módosul, hogy a hozzákötődési valószínűség itt már a fokszám és a η fitneszérték szorzatával arányos:

$$P(\text{az } i\text{-t választjuk}) = \frac{k_i \cdot \eta_i}{\sum_j (k_j \eta_j)} \qquad 63.$$

[34] DOROGOVTSEV–MENDES 2003.

[35] BARABÁSI et al. 2000; BIANCONI–BARABÁSI 2001.

Ebben a modellben a kialakuló hálózat viselkedése a fitnesz változó eloszlásától függ, amelynek megfelelő megválasztásával elérhetjük azt, hogy a fittebb, de később megjelenő csomópontok idővel az öregebb, de kevésbé fitt pontok fölé tudjanak kerekedni (49. ábra).



49. ábra

Különböző időpontokban csatlakozó csomópontok fokszámának időfejlődése

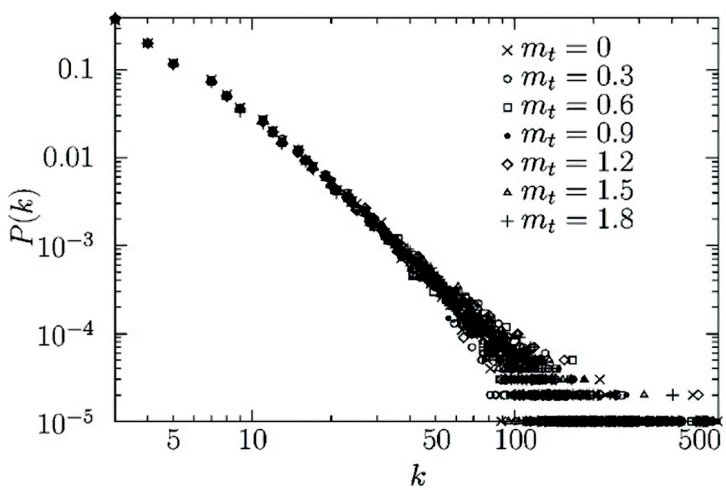
Megjegyzés: A felső sorban az eredeti Barabási–Albert-modellben láthatjuk a $t_i = 1000, 3000, 5000$ időlépésben megjelenő csomópontok fokszámának alakulását lineáris és logaritmikus skálán. Az alsó sor már a Bianconi–Barabási-modellre vonatkozik, ahol a csomópont fitneszének is van hatása a fokszámra, hiszen itt láthatóan „beelőz" egy később csatlakozó csomópont.

Forrás: Barabási 2017

- *Magas klaszterezettségű skálafüggetlen modellek.* Az eredeti Barabási–Albert-modell nagy rendszerméreteknél nem adja vissza a valós hálózatokban megfigyelhető magas klaszterezettségi együtthatót. Ennek kiküszöbölésére születtek olyan modellvariációk, amelyek a kapcsolódási mechanizmus minimális kiegészítésével lehetővé teszik, hogy a létrejövő skálafüggetlen hálózatban a $\langle C \rangle$ értékét is állítani lehessen a paraméterek változtatásával. A Petter Holme és Beom Jun Kim által javasolt megközelítésben az új csomópontok bekötési szabálya a következő:[36]

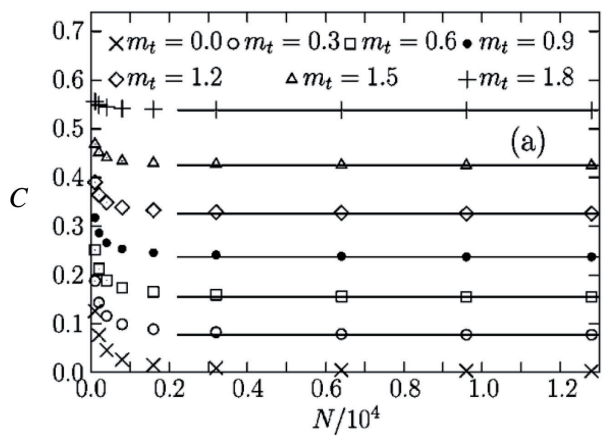
[36] Holme–Kim 2002.

  - Az új csomópont a fokszámmal arányos preferencia alapján választja ki magának az első szomszédját a már létező csomópontok közül úgy, mint az eredeti Barabási–Albert-modellben. Ez biztosítja azt, hogy a létrejövő hálózat skálafüggetlen (50. ábra).
  - Ha $m > 1$, akkor az összes további $m - 1$ új él bekötésénél p valószínűséggel az előzőleg választott régi csúcs egy szomszédjához kapcsolódik, illetve $(1 - p)$ valószínűséggel pedig megint preferenciálisan választ. (A p egy rögzített paraméter.)
- Látható, hogy amennyiben az előzőleg választott régi csúcs szomszédjához kapcsolódunk, úgy garantáltan létrehozunk egy háromszöget, ami természetesen növeli a hálózat átlagos klaszterezettségi együtthatóját. Ennek révén a p paraméter segítségével lehet a kapott hálózat $\langle C \rangle$ értékét hangolni (51. ábra), és nyílik meg annak lehetősége, hogy reprodukáljuk a valós hálózatokban mérhető, viszonylag magas átlagos klaszterezettséget.



50. ábra
A Holme–Kim-modell fokszámeloszlása

Megjegyzés: Az m_t paraméter a csúcsonként várható „háromszögformáló” élek számát jelöli, $m_t = (m - 1)p$.
Forrás: HOLME–KIM 2002



51. ábra

A Holme–Kim-modell átlagos klaszterezettségi együtthatója

Megjegyzés: A szimulációk során mért $\langle C \rangle$ a rendszerméret függvényében lett feltüntetve, és látható, hogy adott m_t esetén egy jól meghatározható értékhez konvergál.

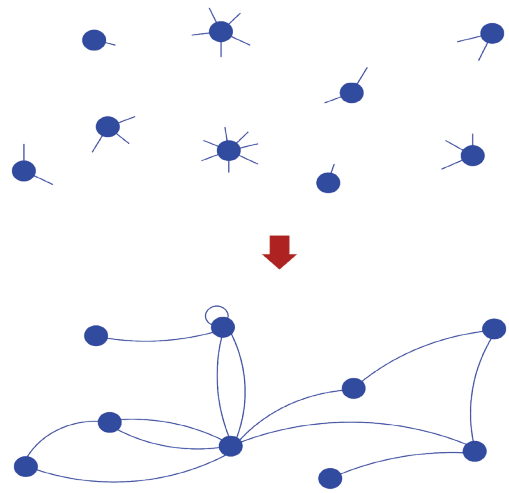
Forrás: Holme–Kim 2002

Konfigurációs modell

Valós hálózatok elemzésekor sokszor merül fel, hogy érdemes lenne az eredményeket összehasonlítani egy olyan véletlen gráffal, amelynek ugyanaz a fokszámeloszlása, viszont a kapcsolatok (ezen megkötés mellett) teljesen véletlenszerűek. Korábban, például a *Strukturális diszasszortativitás* című alfejezetben megállapítottuk, hogy skálafüggetlen diszasszortatív hálózatok esetén érdemes a fokszám-korrelációt lemérni egy megegyező fokszámeloszlású, de véletlenszerűen huzalozott hálózatban is annak eldöntésére, hogy a tapasztalt diszasszortatív viselkedés vajon pusztán az extrém inhomogén fokszámeloszlás következménye, vagy esetleg ennél többről van szó. A fentiek alapján elmondható, hogy gyakorlati szempontból teljesen megalapozott az igény egy olyan véletlengráf-modell iránt, amellyel rögzített fokszámeloszlással rendelkező, de amúgy teljesen véletlenszerű kapcsolatokból felépülő hálózatot lehet előállítani. Ezt az igényt elégíti ki a *konfigurációs modell*.[37]

Matematikai definícióját talán úgy a legegyszerűbb megadni, hogy a rögzített N méretű és $p(k)$ fokszámeloszlású gráfok sokaságából indulunk ki, amely tartalmazza az összes lehetséges olyan gráfot, amely N csomópontból áll, és a csomópontok fokszámeloszlása $p(k)$. A konfigurációs modellben ezen sokaság minden elemét egyformán valószínűnek tesszük fel, ezért ebben a modellben egy véletlen gráf előállítása megfelel az összes lehetséges [adott N-hez és $p(k)$-hoz tartozó] hálózat közül egy egyenletes véletlenszerű mintavételnek.

[37] Bollobás 1980; Molloy–Reed 1995; Newman 2010.



52. ábra
A konfigurációs modell

Megjegyzés: Az első lépésben minden csomópont számára kiosztjuk a fokszámot az adott $p(k)$ fokszámeloszlás szerint (felső ábra). Ezután a csomópontokhoz rögzített fél éleket véletlenszerűen kapcsoljuk össze egymással, létrehozva egy $p(k)$ fokszámeloszlású véletlen gráfot. Az összekapcsolás során előfordulhat, hogy többszörös élek, illetve mindkét végükkel azonos csomóponthoz kapcsolódó élek is keletkeznek.

Forrás: a szerző szerkesztése

Lévén, hogy az adott $p(k)$ fokszámeloszláshoz tartozó összes lehetséges hálózatok száma N függvényében extrém gyorsan emelkedik, a konfigurációs modell gyakorlati megvalósításakor az előre legyártott lehetséges hálózatok közti mintavételezés helyett inkább a következő módon járunk el:

- A rögzített $p(k)$ fokszámeloszlás szerint húzunk N darab véletlen számot, ezek lesznek a csomópontok fokszámai.
- Képzeletben minden csomóponthoz csatlakoztatunk a fokszámával egyező számú fél élt, amelyek egyelőre nem csatlakoznak semelyik másik ponthoz sem (52. ábra).
- Majd a különböző csomópontokhoz rögzített fél élek szabad végeit teljesen véletlenszerűen elkezdjük egymáshoz kapcsolni, létrehozva ezzel egy $p(k)$ fokszámeloszlású véletlen hálózatot (52. ábra).

Az a gondolat, hogy a hálózatot adott fokszámeloszlású, de kezdetben egymástól izolált, csak fél élekkel rendelkező csomópontok véletlenszerű összekötögetésével hozzuk létre, már szerepelt a fokszám-korrelációk tárgyalásánál, ahol például levezettük, hogy annak valószínűsége, hogy egy k_i és egy k_j fokszámú csomópont összekapcsolódik, $k_i k_j / 2L$ (a 40. egyenlet szerint). Érdemes ezzel kapcsolatban megjegyezni, hogy skálafüggetlen $p(k)$ esetén az extrém nagy fokszámú csomópontok között várható élek száma könnyen 1 fölé nőhet, ami többszörös élek, valamint mindkét végükkel egyazon csúcshoz csatlakozó

élek megjelenéséhez vezethet. (Ezt illusztrálja a 52. ábra, és ez az effektus okozza továbbá a strukturális diszasszortativitást).

Amennyiben eleve egy pszeudográfot szeretnénk létrehozni, ez a jelenség nem jelent gondot, ugyanakkor a társadalom és a természet komplex rendszereit leíró hálózatok esetén az a bevett szokás, hogy egyszerű gráfokban gondolkodunk. Annak érdekében, hogy a fél élek összekapcsolása ne vezessen többszörös kapcsolatok vagy önkapcsolatok megjelenéséhez, a következő módosításokkal lehet élni a fent ismertetett eljárás során:

- Minden összekapcsolás előtt ellenőrizzük, hogy a kiválasztott két fél él összekötése nem vezet-e tiltott él megjelenéséhez. Amennyiben igen, az adott összekapcsolást elvetjük, és a két fél él visszakerül a még nem párosított fél élek közé.
- Érdemes a legnagyobb fokszámú csúcsokhoz rögzített fél élekkel kezdeni az összekötögetés folyamatát, nehogy úgy járjunk, hogy már csak olyan szabad fél élek maradtak a rendszerben, amelyek összekötése mindenféleképpen tiltott élek megjelenéséhez vezetne.

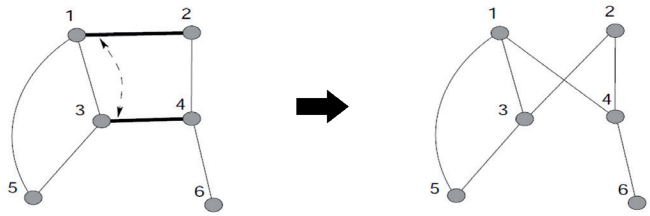
Amennyiben a fokszám sorrendjében haladunk végig a csomópontokon, és keresünk párt a hozzájuk rögzített fél éleknek, jó eséllyel sikerülhet az adott $p(k)$ fokszámeloszlással rendelkező egyszerű gráfot előállítani. Azonban ezzel az a gond, hogy pontosan a rögzített sorrend miatt nem tekinthetünk rá úgy, mint egy egyenletesen véletlenszerűen választott elemre az összes lehetséges $p(k)$ fokszámeloszlású (és N méretű) hálózat közül. Emiatt érdemes még az összekapcsolási folyamat után a hálózaton további véletlenszerű élátkötéseket is végezni úgy, hogy a csomópontok fokszámát megőrizzük. Lévén, hogy ez a fajta randomizálás a konfigurációs modell keretein kívül is fontos, egy külön alfejezetben foglalkozunk vele.

Fokszámmegtartó randomizálás

A fokszámmegtartó randomizálás célja az, hogy a hálózatban véletlenszerű élátkötéseket hajtsunk végre úgy, hogy közben egyik csúcs fokszáma sem változik meg.[38] Ennek alaplépése általában az, hogy két él egy-egy végpontját felcseréljük egymással, így a fokszámok változatlanok maradnak, viszont a szomszédsági viszonyok megváltoznak. Két fajta implementációt szokás megkülönböztetni:

- *Élrandomizálás:* Minden egyes iterációnál az élek listájából egyenletesen véletlenszerűen választunk két élt, amelyek között egy-egy végpontot kölcsönösen kicserélünk (52. ábra). Természetesen, ha egyszerű gráfhoz ragaszkodunk, akkor a felcserélés előtt ellenőrizni kell, hogy nem hozunk-e létre tiltott éleket a cserével (és amennyiben igen, az adott cserét nem hajtjuk végre).

[38] Maslov–Sneppen 2002.

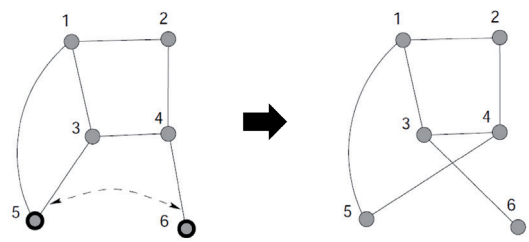


53. ábra
Élrandomizálás

Megjegyzés: Egy lépésben a kapcsolatok listájából választunk két élt véletlenszerűen, majd egyik végpontjukat felcseréljük.

Forrás: a szerző szerkesztése

- *Csúcsrandomizálás:* Egy elemi lépés során itt a csomópontok listájából választunk egyenletesen véletlenszerűen két pontot, amelyek között kölcsönösen egy-egy élt kicserélünk (54. ábra). Hasonlóan az élrandomizáláshoz, ha egy egyszerű gráfból indulva nem szeretnénk többszörös élek és önkapcsolatok megjelenését engedélyezni, akkor a csere végrehajtása előtt egy ellenőrző lépést kell beiktatni.



54. ábra
Csúcsrandomizálás

Megjegyzés: Iterációnként a csomópontok közül választunk véletlenszerűen egy párt, és köztük egy-egy élt kicserélünk.

Forrás: a szerző szerkesztése

A gyakorlatban az élrandomizálás a népszerűbb, mert ennél ha mondjuk $c \cdot L$ iteráción keresztül folytattuk a véletlenszerű átkötéseket, akkor elmondhatjuk, hogy minden élre átlagosan c véletlen átkötés jutott.

A fokszámmegtartó randomizáló algoritmusok nagy haszna, hogy nagyon egyszerűvé teszik a valós hálózatok és a konfigurációs modell közötti összehasonlítást: ahhoz, hogy megkapjuk egy valós hálózat konfigurációs modellen belüli megfelelőjét, nem kell mást tennünk, mint randomizálni az eredeti hálózatot fokszámmegtartó módon. Ezzel kihagyhatjuk a fokszámeloszlás alapján történő véletlenszerű fokszámkiosztást, és a már rögzített fokszámú, de még egymástól izolált csomópontok összekötését végző lépéseket.

Egy valós hálózat elemzése során a fokszámmegtartó randomizálás útján előálló véletlen gráf elméleti szempontból is nagyon fontos: megmutatja, hogy milyen lenne a vizsgált rendszer akkor, ha a fokszámeloszláson kívül semmilyen megkötés nem vonatkozna a kapcsolatok kialakulására. Ez nagyon fontos volt például a strukturális diszasszortativitás vizsgálatakor, ahol az volt a kérdés, vajon pusztán a fokszámeloszlás skálafüggetlen természete okozza-e a fokszám-korreláció diszasszortatív viselkedését, vagy esetleg van ezen felül valami egyéb hatás is, amely diszasszortatív irányba módosítja a fokszám-korrelációt. Ennek eldöntéséhez a fokszám-korrelációt a hálózat randomizált másolataiban is le kell mérni, és a kapott görbéket össze kell vetni az eredeti hálózat $k_{\mathrm{nn}}(k)$ görbéjével: ha a grafikonok nagyjából egybeesnek, akkor lényegében pusztán a $p(k)$ alapján magyarázható a fokszám-korreláció, míg szignifikáns eltérés esetén joggal feltételezhetjük, hogy valami egyéb jelentős hatás is befolyással van a $k_{\mathrm{nn}}(k)$ alakjára.

Felhasznált irodalom

Albert, R. – Jeong, H. – Barabási A.-L. (2000): Error and attack tolerance of complex networks. *Nature,* No. 406. 378–382. DOI: https://doi.org/10.1038/35019019

Arquilla, J. – Ronfeldt, D. (2001): *Networks and Netwars: The Future of Terror, Crime, and Militancy.* Santa Monica, CA, RAND Corporation.

Backstrom, L. – Boldi, P. – Rosa, M. – Ugander, J. – Vigna, S. (2012): Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference.* New York, ACM Press. 33–42.

Barabási, A.-L. – Albert, R. (1999): Emergence of scaling in random networks. *Science,* Vol. 286, No. 5439. 509–512. DOI: https://doi.org/10.1126/science.286.5439.509

Barabási, A.-L. – Albert, R. – Jeong, H. – Bianconi, G. (2000): Power-law distribution of the world wide web. *Science,* Vol. 287, No. 5461. 2115. DOI: https://doi.org/10.1126/science.287.5461.2115a

Barabási A.-L. (2017): *A hálózatok tudománya*. Budapest, Libri Könyvkiadó.

Bianconi, G. – Barabási A.-L. (2001): Competition and multiscaling in evolving networks. *Europhysics Letters,* Vol. 54, No. 4. 436–442. DOI: https://doi.org/10.1209%2Fepl%2Fi2001-00260-6

Bollobás, B. – Riordan, O. (2004): The diameter of a scale-free random graph. *Combinatorica,* Vol. 24, No. 1. 5–34. DOI: https://doi.org/10.1007/s00493-004-0002-2

Bollobás, B. (1980): A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.,* Vol. 1, No. 4. 311–316. DOI: https://doi.org/10.1016/S0195-6698(80)80030-8

Cohen, R. – Havlin, S. (2003): Scale-free networks are ultrasmall. *Phys. Rev. Lett.,* Vol. 90, No. 5. DOI: https://doi.org/10.1103/PhysRevLett.90.058701

Dorogovtsev, S. N. – Mendes, J. F. F. (2003): *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford, Oxford University Press.

Erdős P. – Rényi A. (1959): On random graphs I. *Publicationes Mathematicae (Debrecen),* Vol. 6, No. 3–4. 290–297.

Erdős P. – Rényi A. (1960): *On the evolution of random graphs*. Különnyomat a Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményeiből. Budapest, Akadémiai Kiadó. 17–61.

Freeman, L. C. – Thompson, C. R. (1989): Estimating acquaintanceship volume. In Kochen, M. ed: *The Small World*. Norwood, N.J., Ablex. 147–158.

GOH, K.-I. – CUSICK, M. E. – VALLE, D. – CHILDS, B. – VIDAL, M. – BARABÁSI A.-L. (2007): The human disease network. *Proc. Natl. Acad. Sci. USA,* Vol. 104, No. 21. 8685–8690. DOI: https://doi.org/10.1073/pnas.0701361104

HOLME, P. – KIM, B. J. (2002): Growing scale-free networks with tunable clustering. *Phys. Rev. E,* Vol. 65, No. 2. DOI: https://doi.org/10.1103/PhysRevE.65.026107

JEONG, H. – ALBERT, R. – BARABÁSI A.-L. (1999): Diameter of the world-wide web. *Nature,* Vol. 401. 130–131. DOI: https://doi.org/10.1038/43601.

JEONG, H. – MASON, S. P. – BARABÁSI, A.-L. – OLTVAI, Z. N. (2001): Lethality and centrality in protein networks. *Nature,* Vol. 411. 41–42. DOI: https://doi.org/10.1038/35075138

KARINTHY F. (1929): Láncszemek. In KARINTHY F.: *Minden másképp van*. Budapest, Atheneum Irodalmi és Nyomdai R.-T. 85–90.

KLEINBERG, J. M. – KUMAR, R. – RAGHAVAN, P. – RAJAGOPALAN, S. – TOMKINS, A. S. (1999): The web as a graph: measurements, models, and methods. In ASANO, T. – IMAI, H. – LEE, D. T. – NAKANO, S. – TOKUYAMA, T. eds.: *Computing and Combinatorics*. Berlin–Heidelberg, Springer. 1–17. DOI: https://doi.org/10.1007/3-540-48686-0_1

KUMAR, R. – RAGHAVAN, P. – RAJALOPAGAN, S. – DIVAKUMAR, D. – TOMKINS, A. S. – UPFAL, E. (2000): The web as a graph. In *Proceedings of the Nineteenth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems.* New York, ACM. 1–10. DOI: https://doi.org/10.1145/335168.335170

MASLOV, S. – SNEPPEN, K. (2002): Specificity and stability in topology of protein networks. *Science,* Vol. 296, No. 5569. 910–913. DOI: https://doi.org/10.1126/science.1065103

MILGRAM, S. (1967): The small world problem. *Psychology Today,* Vol. 1, No. 1. 61–67.

MOLLOY, M. – REED, B. A. (1995): Critical point for random graphs with a given degree sequence. *Random Structures & Algorithms,* Vol. 6, No. 2–3. 161–180. DOI: https://doi.org/10.1002/rsa.3240060204

NEWMAN, M. E. J. (2010): *Networks: An Introduction*. Oxford, Oxford University Press.

PALLA G. – DERÉNYI Imre – FARKAS I. – VICSEK T. (2005): Uncovering the overlapping community structure of complex networks in nature and society. *Nature,* Vol. 435. 814–818. DOI: https://doi.org/10.1038/nature03607

PASTOR-SATORRAS, R. – VÁZQUEZ, A. – VESPIGNANI, A. (2001): Dynamical and correlation properties of the internet. *Phys. Rev. Lett.,* Vol. 87, No. 25. DOI: https://doi.org/10.1103/PhysRevLett.87.258701

PASTOR-SATORRAS, R. – VESPIGNANI, A. (2001): Epidemic spreading in scale-free networks. *Phys. Rev. Lett.,* Vol. 86, No. 14. 3200–3203. DOI: https://doi.org/10.1103/PhysRevLett.86.3200

VÁZQUEZ, A. – PASTOR-SATORRAS, R. – VESPIGNANI, A. (2002): Large-scale topological and dynamical properties of the internet. *Phys. Rev. E,* Vol. 65, No. 6. DOI: https://doi.org/10.1103/PhysRevE.65.066130

WATTS, D. J. – STROGATZ, S. H. (1998): Collective dynamics of 'small-world' networks. *Nature,* Vol. 393. 440–442. DOI: https://doi.org/10.1038/30918

WILSON, C. (2010): *Searching for Saddam: A five-part series on how the U.S. military used social networking to capture the Iraqi dictator.* Elérhető: www.slate.com/id/2245228/ (A letöltés dátuma: 2019. 03. 12.)

Bányász Péter – Dobos László – Palla Gergely – Pollner Péter

Lélektani műveletek a közösségi médiában

Bevezetés

A közösségimédia-használat napjaink megkerülhetetlen jelensége. A kezdetekben a közösségi oldalak a kapcsolattartás eszközeként jelentek meg, azonban az évek során életünk számos területén váltak meghatározóvá. A közösségi média az állandó változás terepe, ez azonban a természethez hasonló érzékeny ökoszisztémát hozott létre.[1] A közösségi oldalak alapvetően az online hirdetésekből realizálják bevételüket, amelynek növelése érdekében minél nagyobb számú, elkötelezett felhasználóval kell, hogy rendelkezzenek. Annak érdekében, hogy bővítsék a felhasználók számát, illetve hogy megtartsák a már meglevő felhasználókat, a közösségi oldalak állandó innovációra kényszerülnek, legyen szó új szolgáltatások bevezetéséről,[2] a riválisok szolgáltatásának adaptálásáról[3] vagy a feltörekvő riválisok felvásárlásáról és magukba integrálásáról.[4] Az érzékeny ökoszisztémát napjaink legnépszerűbb közösségi oldala, a Facebook szemlélteti leginkább.

Mint az köztudott, a Facebook a hírfolyamának állandó alakítgatásával igyekszik megtartani a felhasználóit. Ennek oka egyszerű: minél több ismerőse van valakinek a Facebookon, vélhetően annál több számára irreleváns, érdektelen tartalommal találkozik a saját hírfolyamában. Annak érdekében, hogy a felhasználókat ne veszítsék el, különböző beavatkozásokat kísérelnek meg, letilthatunk számunkra érdektelen embereket, tartalomtípusokat, az algoritmus különböző tartalmakat preferál, másokat háttérbe szorít.[5] Ennek egyik legfontosabb eszköze a Facebook algoritmusa, amelynek pontos működéséről nincs információnk, csak bizonyos szempontokat ismerünk. Az algoritmus, felmérve a felhasználó preferenciáit, igyekszik kiszolgálni az elemzés során megállapított érdeklődési körét.

---

[1] A természetben gyakran egy minimális beavatkozás is olyan események láncolatát indítja el, amely a környezet teljes megváltozásával jár. Erre kiváló példával szolgál az amerikai Yellowstone Nemzeti Park, ahol az 1930-as években kiirtották a szürke farkasokat, aminek következtében jelentősen megnőtt a jávorszarvas-populáció, ez pedig katasztrofális hatással bírt a park ökoszisztémájára. A változást 1995-ben a farkasok újrabetelepítése hozta el, amely az elszaporodott jávorszarvasok és prérifarkasok természetes szabályozásával a haldokló flórát és faunát megmentette és növekedésnek indította. Bővebben lásd: Peglar 2018.

[2] Lásd például: videochat bevezetése, trending topics megjelenése, a hírfolyam módosítása az algoritmus által stb.

[3] Erre szolgál példaként a Snapchat mintájára bevezetett „Napom" funkció.

[4] Lásd például az Instagram képmegosztó vagy WhatsApp üzenetküldő szolgáltatást.

[5] Egy időben népszerűek voltak a kattintásvadász címek, mint például: „Nem fogod elhinni, mit tett ez a kislány a közönség előtt. Mindenkinek leesett az álla" stb. Ezeket a Facebook sokáig támogatta, az ilyen címet viselő tartalmakat nagy számban jelenítette meg a felhasználók hírfolyamában, de az álhírek elleni harc jegyében száműzte ezek láthatóságát.

Ennek érdekében több ezer szempont alapján értékeli a Facebook a felhasználókat, ami egyúttal a hirdetések megjelenítésében is kiemelt jelentőségű. Ennek érdekében a Facebook algoritmusa többek között figyeli, kikkel beszélünk gyakran, milyen témákban, milyen tartalmakat fogyasztunk, kiknek a posztjait kommenteljük stb. Ezek (és egyéb variánsok) alapján az algoritmus azokat a tartalmakat jeleníti meg előttünk, amelyeket értékelése alapján érdekesnek, relevánsnak ítélünk.

A Facebook algoritmusának működése, illetve az apró finomhangolások azonban a természetes ökoszisztémához hasonlóan számos területen eredményeztek radikális változásokat a spill over hatás következtében. Az egyik ilyen terület a politikai döntéshozatal befolyásolásában azonosítható, amelyet napjaink kiemelt biztonsági kockázataként nevesítenek. Jelen tanulmány ennek rendszerszintű vizsgálatát látja el egy, a 2016-os amerikai elnökválasztási kampányból vett álhír bemutatásával. Az álhírek terjedésében számos ponton azonosíthatunk hálózatokat. A közösségi hálózatok egyrészt kiemelt területei az álhírek megosztásának, de ezenfelül a terjesztés módszerei kapcsán is találkozhatunk olyan botnethálózatokkal, amelyek segítik az álhírek minél szélesebb körben történő megismerését. Izgalmas kérdés azoknak a szereplőknek az azonosítása, akik az álhírek terjesztését végzik. A különböző oldalak, csoportok, szervezetek hálózatba rendezése segít megérteni, hogy bizonyos témák hogyan befolyásolják a döntéshozatalt, milyen mintázat alapján terjednek bizonyos álhírek, illetve segíthet a terjesztés mögötti érdekek felismerésében.

A lélektani műveletek helye és szerepe

Lélektani műveletek alatt azokat a tevékenységeket értjük, amelyek során a szemben álló felek céljaik megvalósítása érdekében tudatos lélektani ráhatást alkalmaznak.[6] A lélektani műveleteket az információs műveletek tevékenységi körébe sorolhatjuk.[7] Céljuk egyrészt az információs fölény kialakítása, ezáltal pedig a vezetési fölény megszerzése,[8] másrészt pedig időcsökkentést elérni a saját oldali vezetési folyamat számára, ezzel párhuzamosan az ellenfél oldalán növelni az időt.

Az információs műveleteket egyaránt alkalmazzák támadó vagy védelmi célból annak érdekében, hogy hatást váltsanak ki a katonai, gazdasági, politikai alrendszerben. A támadó jellegű információs művelet esetében a cél az, hogy a speciális érdekekre vagy speciális fenyegetésekre választ adva gyakoroljanak hatást az ellenfélre akár békében, válságban

[6] PIX 2005.

[7] Idetartozik továbbá az elektronikai hadviselés, a dezinformáció, az információs célpontok fizikai pusztítása, a műveleti biztonság, valamint a számítógép-hálózati műveletek. A témáról bővebben lásd: HAIG–VÁRHEGYI 2005.

[8] Haig Zsolt megfogalmazása alapján: „Az információs fölény a két szembenálló fél közötti relatív viszonyt jelenti, amely felhasználható a saját célok, érdekek másik félnél eredményesebb érvényesítésére. […] a végcél a vezetési folyamatban jelentkező vezetési fölény elérése. Az információs fölény alapvető funkciója tehát, hogy kedvező információs helyzetet, tudástöbbletet teremtsen a vezetési fölény kialakításához. A vezetési fölény egyrészről a szembenálló felek vezetési folyamatai között minőségi különbséget jelent: az egyik fél tevékenységét meghatározó intézkedések, utasítások tartalma és időbelisége lényegesen jobban tükrözi a kialakult helyzetet és az ahhoz alkalmazódó célszerű cselekvésmodellt, mint a másiké. Másrészről azt az állapotot fejezi ki, amikor ugyanezen fél végrehajtói eltökéltsége az utasítások teljesítésére azonos vagy nagyobb, mint a másik fél (társadalom) tagjaié." Bővebben lásd: HAIG et al. 2014.

vagy konfliktus idején. Ezzel szemben a védelmi információs művelet során a cél az, hogy megvédjék a saját információkat, valamint fenntartsák az információkhoz való hozzáférést, illetve elősegítsék az információs rendszerek hatékony használatát. Az információs műveletek fogalmát korábban információs hadviselésként is használták, civil szóhasználatban mai napig jellemző, azonban a katonai szakterminológia a NATO esetében információs műveletekként nevesíti a tevékenységet. A médiában az információs hadviselés és információs műveletek közti különbséget általában az elérendő célban választják el: az információs hadviselés civil célok (például gazdasági, politikai célok) elérésére szolgál, az információs művelet pedig katonai célok elérésére vonatkozik. Ez azonban nehezen mérhető kategória. Nem könnyen azonosíthatjuk az egyes célokat, hogy vajon civil vagy katonai célok megvalósítására végzik-e a tevékenységet. Gazdasági, politikai nyomásgyakorlás ugyanúgy lehet katonai cél.

Legyen szó katonai vagy civil cél megvalósításáról, a lélektani műveletek mindkettő esetében fontos szerepet tölthetnek be. A Magyar Honvédség által, 2014-ben kiadott *Információs műveletek* doktrína alapján: „A Lélektani Műveletek (PSYOPS) elsődleges célja, hogy befolyásolja egy kiválasztott célcsoport viselkedését, magatartásformáit és véleményét az elöljáró által elfogadott PSYOPS célokkal összhangban, valamint hogy kiváltsa vagy megerősítse a célcsoport kívánt viselkedését az elöljáró távlati céljainak érdekében."[9] A célcsoport nem csupán egy ellenséges ország lakossága lehet, ugyanúgy irányulhat szövetséges vagy semleges országok lakosságának befolyásolására, de a saját lakosság is lehet lélektani műveletek célcsoportja. Maga a tevékenység az emberi történelem kezdetéig visszavezethető a harci cselekedetek terén. A harcosok varázslókkal történő „felszentelése", az ellenség különböző technikákkal történő, az érzékszervekre ható megrettentése (harci dobok, csontok, vörös festék) mind-mind a lélektani műveletek előzményeihez sorolhatók. Az ellenség manipulálását mint követendő stratégiát már Szun Ce is megfogalmazta *A háború művészete* című művében. A lélektani műveletek fogalmát az 1960-as évektől használják katonai terminusként.

A lélektani műveletek tervezése kapcsán három módszert azonosíthatunk:

- A reflexív kontroll módszere, amely az ellenséges erők parancsnokának döntéshozó mechanizmusát igyekszik befolyásolni. Ennek során első lépésként széles körű felderítést alkalmaznak, majd behatolnak a döntéshozó információs rendszereibe, amelyekben olyan módon helyeznek el álinformációkat, hogy azok a támadók számára kívánt döntések meghozatalát támogassák.
- A társadalmi vírus koncepciója szerint egy társadalmat belülről is meg lehet fertőzni, amennyiben megfelelő pozíciókban levő személyek terjesztenek bizonyos álhíreket. Ezek a személyek lehetnek a támadók által beépített egyének vagy olyan véleményvezérek, akiket akár „idegen zászló alatt" beszerveznek a vélemények befolyásolására. Ezek a személyek általában ideológiájukban befolyásolható, politikailag sértett, a világ történéseiről objektíven nem informálódó egyének. A véleményvezéreket, influencereket hálózatok segítségével könnyen azonosítani lehet, ami megkönnyíti a lélektani műveletek tervezését.
- A hagyományos eszközök (például röplap, hangszórós alakulatok) mellett különleges módszereket és eszközöket alkalmaznak, amelyek célja az adott személy

[9] Ált/57, Információs műveletek doktrína. MH DOFT kód: MD 3.10 (1). 2014. 08. 05. 1–15.

valóságérzetének befolyásolása, lelki stabilitásának csökkentése. Napjainkban a különleges eszközök egyik legfontosabb ágát a mesterségesintelligencia-kutatások jelentik. Az úgynevezett „DeepFake"-technológia gépi mélytanulás segítségével lehetővé teszi olyan, akár valós időben közvetített audiovizuális tartalmak közvetítését, amelyekben valakinek az arcát, illetve hangját rámontírozzák egy másik személyre. A közösségi médiában terjedő álhírek jelentős részét bizonyos sémák felismerésével könnyen azonosíthatjuk, azonban egy ilyen esetben még a rutinos, az álhíreket nagy hatékonysággal felismerő személynek is nehéz dolga akad. A technológiai fejlődés vélelmezhetően ezt az irányt tovább fogja erősíteni.

A köznyelvben a lélektani műveleteket gyakran a propagandával azonosítják, azonban a propaganda jelentős szűkítés a lélektani műveletek spektrumához képest, csupán részhalmazként tekinthetünk rá. A propagandát három kategória mentén csoportosítjuk:

- A fehér propaganda jellemzője, hogy ismert, ki közvetíti, gyakran valósághű, a hírforrása hiteles, így a terjedését könnyű megakadályozni, cáfolni a valótlanságokat. Eszköztárának jellegzetes példái a vicclapok, karikatúrák, amelyekkel az ellenséget teszik nevetségessé.
- A fekete propaganda a valóságtól eltérő hírközlést jelent, alkalmazói gyakran álcázzák önmagukat, megtévesztve a célközönséget, mintha az nem az ellenségtől származna, hanem a saját kormánytól.
- A szürke propaganda esetében nem ismert a hír forrása. Fő célja az ellenség demoralizálása olyan hamis, alapvetően az ellenség helyzetéről szóló hírek terjesztésével, amelyekkel csökkentik a harci kedvet, morált. Ide tartozhat például a hátországból származó álhírek terjesztése, amely a katonák családjainak pusztulásáról vagy éppen feleségeik, barátnőjük hűtlenségéről szól.

A NATO lélektani műveli doktrínája[10] a célzott információközlést fogalmazza meg alapvető gyakorlatként, amelynek egyik fő oka az, hogy a propaganda a közvélekedésben rendszerint összemosódik valamilyen politikai ideológiával.[11] A célzott információközlés rendkívül széles skálán mozoghat. A technikai és technológiai innovációval párhuzamosan bővültek az információterjesztés médiumai. Napjainkban az egyik legjelentősebb csatorna az internet és azon belül a közösségi média.[12] A különböző közösségi hálózatok, blogok, fórumok, kép- és videómegosztó oldalak mind lehetőséget biztosítanak propaganda és ellenpropaganda végzésére, amelyre az egyes államok kiterjedt szervezeteket tartanak fenn. Ennek egyik legismertebb eljárása a közösségi oldalakon való álhírek terjesztése, ami a kormányzati szereplők mellett az egyénekre, üzleti szereplőkre is fenyegetést jelent.[13]

A közösségi média ilyen mértékű felértékelése nem véletlen, hiszen a vonatkozó statisztikák szerint rengetegen használják napi szinten a különböző közösségi oldalakat. A 2018-as Digitális Gazdaság és Társadalom Index (DESI) mérése alapján hazánkban a közösségimédia-használat 84%-ra tehető az internetezők körében.[14] Az Európai Unióban ezzel

[10] *AJP-3.7. NATO Military Policy on Psychological Operations* (2003).
[11] Miller 2015.
[12] Briant 2018; Dudatyev 2017.
[13] Traylor–Freese–Wong 2017.
[14] Európai Bizottság 2018.

a második helyen szerepelünk, az uniós átlag csupán 65%. Ez mindenképpen figyelemre méltó eredmény úgy, hogy Magyarország egyéb területeken igen komoly lemaradásban van a DESI mérései alapján. A jelentésből az is kiderül, hogy a hírfogyasztási szokásaink nagymértékben a közösségi médiára kerültek át, ez a magyar internetezők esetében 85%-ra tehető.

Álhírek a közösségi médiában

Álhírek alatt azokat a híreket értjük, amelyek kitaláción alapulnak, semmilyen tény nem támasztja alá ezeket. Az álhíreknek azonban széles spektrumát különböztethetjük meg. Melissa Zimdars gyűjteményét alapul véve az alábbi kategóriákat azonosíthatjuk:

- álhírek;
- elfogult hírek: a megjelenő hírek egy bizonyos politikai oldal, ideológia mellett súlyosan torzított formában, egyoldalúan jelennek meg;
- áltudományos hírek, tudományosan nem alátámasztott, súlyosabb esetben a tudományos eredményeknek ellentmondó hírek, például az oltásellenességgel kapcsolatos hírek;
- összeesküvés-elméletek;
- szatirikus hírek: olyan humorosan megírt hírek, amelyek bár álhírek, céljuk azonban a szórakoztatás;
- kattintásvadász hírek: olyan hírek, amelyek mind címükben, mind tartalmukban félrevezetők, céljuk, hogy minél többen kattintsanak az oldalra, hogy ily módon növeljék a reklámbevételeket.[15]

Az egyes kategóriákat a politikai döntéshozatal befolyásolására is használják, hiszen a kitűzött célt nemcsak akkor lehet elérni, ha az olvasók igaznak tekintenek egy hírt, hanem akkor is, ha más jellegű hírekkel bizonytalanságot keltenek, a hivatalosnak tekinthető híreket megkérdőjelezik.

Az úgynevezett „post-truth”-jelenség erősödése szorosan összefügg azzal, hogy az álhírek hatékonyan képesek befolyásolni a politikai döntéshozatalt.[16] A post-truth egy olyan állapotot ír le, amikor a közvéleményt nem a tények, hanem az érzelmek, a meggyőződésen alapuló hitek határozzák meg. Az objektivitás egyre inkább a háttérbe szorul a valóságról alkotott percepcióval kapcsolatban, helyét számos párhuzamos, szubjektív valóság váltja fel. A párhuzamos valóságok kialakulását a közösségi oldalak működése is támogatja. Azáltal, hogy egy algoritmus megpróbálja kitalálni, milyen típusú tartalom érdekli a felhasználót a szokásainak értékelése-elemzése során, a plurális tartalomfogyasztás hiányában úgynevezett „bubble filter” alakulhat ki. A magyarul szűrőbuboréknak vagy buborékhatásnak nevezett fogalom lényege, hogy a felhasználó csak azokat a tartalmakat, véleményeket találja meg a közösségi oldalakon, amelyeket rendszeresen fogyaszt. Ez azonban oda vezethet, hogy az ellentétes vélemények egyáltalán nem jelennek meg előtte, és megerősíti a felhasználó által vallott, preferált véleményeket. Ily módon a valóságérzékelésében az ellentétes véle-

[15] Zimdars 2016.
[16] Lewandowsky–Ecker–Cook 2017.

mények, még ha egyébként jelentősek is, nem vagy csupán kismértékben jelennek meg, azt a benyomást keltve, mintha a felhasználó beszűkült valóságfelfogása objektív lenne.

A nemzetbiztonsági szolgálatok korán felismerték a „bubble filter"-jelenség politikai döntéshozatal befolyásolásában betöltött szerepét, ami az álhírek célzott terjesztésében is megnyilvánult. A nagy közösségi oldalak, mint a Facebook, bevételeik jelentős részét a személyre szabott hirdetésekből szerzik. Míg korábban a hirdetések célzása csoportok esetében működött, addig napjainkra az egyén pszichológiai jellemzőit is figyelembe veszik. A Cambridge Analytica (CA) nevű cég e tekintetben újítóként értékelhető, és a 2016-os amerikai elnökválasztási kampányban, valamint az azonos évben lezajlott Brexit-népszavazási kampányban is fontos szerepet töltött be.[17] A cég önmagát big data elemző politikai kampánytanácsadóként pozicionálta, és tevékenysége során számos lélektani műveletet is alkalmaz.

A 2012-ben megalapított CA „viselkedésalapú kommunikációs" kampánytanácsadóként hirdette szolgáltatásait, amelyek alapját Aleksandr Kogan, a Cambridge-i Egyetem kutatójának munkássága jelentette. Kogan megalkotta a *This is your digital life* (vagyis *Ez a digitális életed*) nevet viselő Facebook-alkalmazását, amely egy személyiségteszt volt. A Facebook akkori adatvédelmi gyakorlata lehetővé tette az ilyen alkalmazások fejlesztését, azonban meghatározta, kik férhetnek hozzá az így gyűjtött adatokhoz. Az alkalmazás tudatta a felhasználókkal, hogy adatokat gyűjtenek róluk, a kitöltésért még fizetett is. Végül körülbelül 270 ezer felhasználó töltötte ki a személyiségtesztet, de az alkalmazás nemcsak a kitöltő adatait gyűjtötte, hanem az ismerőseinek adataihoz is ugyanúgy hozzáfért.[18] Pontosan nem tudjuk, összesen hány felhasználó adatait gyűjtötte össze az alkalmazás, ugyanis a fokozatosan napvilágra kerülő információk hatására egyre bővült ennek a száma, a kezdeti 50 millióról először 87 millióra. Brittany Kaiser, a CA egykori vezetője a brit parlament bizottsági meghallgatásán azonban azt a vallomást tette, hogy ennél jelentősen több lehet az adatvédelmi botrányban érintett felhasználók száma. 2015-ben a Facebook felfedezte, hogy Kogan harmadik fél részére továbbadta az általa gyűjtött adatokat, ekkor eltávolította az alkalmazást az oldalról, valamint kötelezte Kogant, hogy a begyűjtött adatokat semmisítse meg. Azonban több arra utaló jel is van, hogy bár Kogan igazolta, hogy minden eszközről törölte az adatokat, ez azonban mégsem következett be. A Facebookot ért vádak elsősorban arra vonatkoznak, hogy elhallgatta az esetet 2015-ben, és nem tájékoztatta a felhasználókat, hogy az adataikat nem jogszerűen kezelik.

Az Aleksandr Kogan által kidolgozott személyiségteszt az úgynevezett Big Five személyiségmodelljén alapult, amely szerint bizonyos jellemzők alapján, faktoranalízis segítségével öt különböző faktorcsoportba sorolhatók az egyének: extraverzió, barátságosság, lelkiismeretesség, érzelmi stabilitás, kultúra/intellektus.[19] A Facebook-aktivitás és az egyén személyisége közötti kapcsolatról számos kutatás látott napvilágot, amelyek azt igazolták, hogy a Big Five személyiségmodell alkalmazásával kimutatható az egyes lájkok személyiségünkkel összefüggő kapcsolata.[20] Ha az adatok alapján sikerül megrajzolni az egyén profilját, akkor a targetálás segítségével pontosan olyan hirdetést lehet megjeleníteni számára,

[17] The Cambridge Analytica Files é. n.

[18] Meg kell jegyezni, ez a gyakorlat más alkalmazások esetében is jellemző. A témáról bővebben lásd Symeonidis és szerzőtársai kutatását: Symeonidis et al. 2018.

[19] Norman 1963.

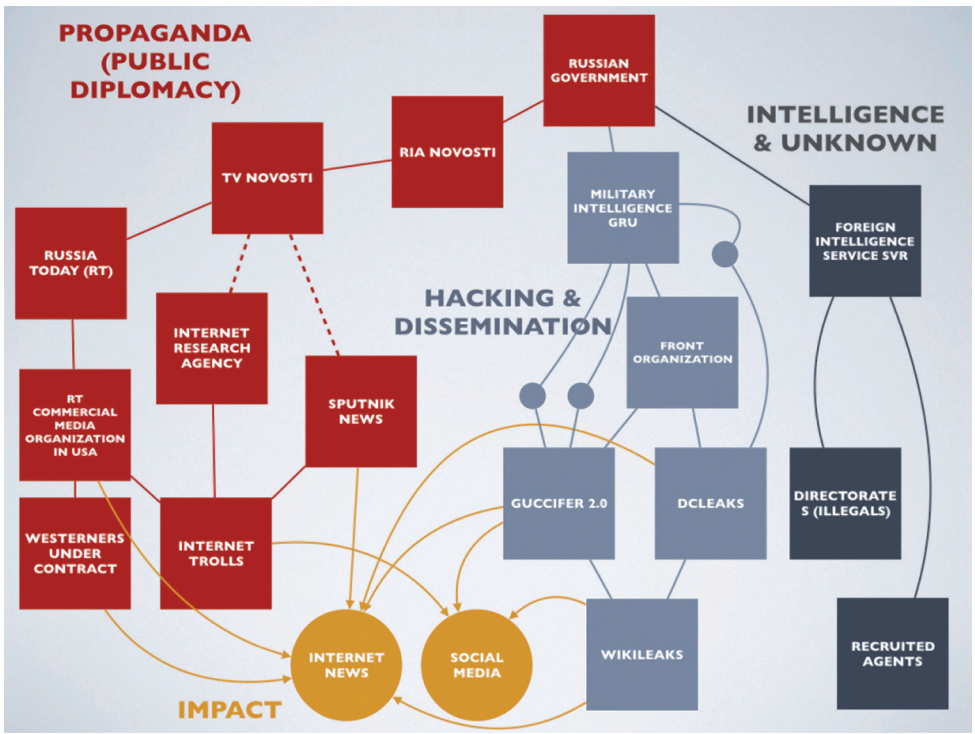
[20] He et al. 2014; Eşkisu–Hoşoğlu–Rasmussen 2017; Azucar–Marengo–Settanni 2018.

amely nagyobb valószínűséggel segíti elő a hirdetésre való kattintást és a vásárlást. Ez az eljárás a választási kampányokban kiválóan működik. Amennyiben a felhasználót egy párt potenciális szavazójának értékel az algoritmus, a személyiségének megfelelő hirdetést jelenít meg számára. Amennyiben neurotikus személyiségzavarra utaló személyiséget állapít meg a felhasználónál, olyan típusú tartalmakat jelenít meg, amelyek erőszakos cselekményekkel függenek össze, míg ha családcentrikus az egyén, akkor inkább a családi hagyományokkal, értékekkel összefüggő hirdetéseket lát.

A 2016-os amerikai elnökválasztás óta Oroszországot rendszeresen éri az a vád, hogy a kiberteret, de azon belül a közösségi médiát aktívan használja idegen államok belpolitikai döntéshozatalának befolyásolására, aminek egyik fő eszköze az álhírek terjesztése. Ennek hatására jelentősen bővült a témával foglalkozó szakirodalom száma.[21] Az orosz lélektani műveletek felerősödése a 2014-es ukrajnai konfliktus kialakulásához vezethető vissza, de aktívan 2008-tól, a grúziai háborútól használják.[22] A kibontakozó orosz–ukrán konfliktusra a szakirodalom hibrid háborúként hivatkozik. Hoffmann fogalmi meghatározását kölcsönözve: „A hibrid fenyegetések a hadviselés számos formáját magukban foglalják, beleértve a konvencionális képességeket, irreguláris harceljárásokat és képződményeket, valamint a válogatás nélküli erőszakot alkalmazó terrorista akciókat és bűnözői tevékenységeket. Hibrid háborúkat egyaránt folytathatnak állami és a legkülönfélébb nem állami szereplők. Az egymástól elszigetelten működő egységek vagy akár ugyanaz a csoport is folytathat »multimodális« tevékenységeket, de ezek általános, műveleti, valamint harcászati irányítása és koordinálása a fő hadszíntéren megy végbe, annak érdekében, hogy a szinergikus hatások bekövetkezzenek a konfliktusok pszichológiai és fizikai dimenzióiban. Ezen hatások a háború valamennyi szintjén jelentkezhetnek.”[23] Nem nehéz belátni, hogy a közösségi média az ott folytatott hírszerzéssel, az álhírekkel kapcsolatos lélektani műveletekkel a hibrid hadviselés fontos elemét képezi.

Az amerikai nemzetbiztonsági szolgálatok nyilvánosságra hoztak egy jelentést, amely különböző szereplőket azonosít az orosz lélektani műveletek végrehajtóiként. Az 1. számú ábrán Edward M. Roche foglalta össze ezeket az aktorokat, akik között hírszerző szolgálatokat, hackereket, illetve a propagandáért felelős entitásokat találhatunk. Ez alapján három részre oszthatjuk a szereplőket: a pirossal jelölt, propagandában érintetteket, idesorolva többek között az internetes trollokat, az Internet Reseach Agencyt, az RT-t, a Sputnik Newst, világosabb kékkel jelölve a hackereket és a terjesztésért felelős csatornákat, ideértve többek között a WikiLeakset vagy a katonai hírszerzést, továbbá sötétebb kékkel jelölve a hírszerző szolgálatokat. Az ábrán sárgával látható az egyes szereplők hatása a közösségi oldalakon, illetve a híroldalakon.

[21] Bővebben lásd például: Figueira–Oliveira 2017; Jang–Kim 2018.
[22] Paul–Matthews 2016.
[23] Hoffman 2007, 8.



1. ábra

Az orosz lélektani műveletek szereplői

Forrás: ROCHE 2017

A RAND Corporation az orosz propaganda sikerét négy tényezőre vezeti vissza:

- számos csatornán közvetítik;
- gyors, ismétlődő formában terjesztik;
- hiányzik belőle az objektivitás;
- nem következetesek.

Az álhírek terjesztését valós személyek és botnethálózatok végezhetik. A botnetek olyan algoritmusok, amelyek különböző álprofilokat hoznak létre a közösségi oldalakon, és ezeken terjesztik az egyes tartalmakat. E botnethálózatok fejlettsége eltérő lehet. 2011-ben kapott nagy nyilvánosságot a US Air Force által kiírt pályázat, amely egy „online identitásmenedzselő szoftverre" vonatkozott.[24] A szoftver egy olyan botnethálózat lett volna, amelyet a közösségi médiában létrehozott álprofilok segítségével a politikai döntéshozatal befolyásolására alkalmaztak volna. Értelemszerűen a szoftvernek meg kellett felelnie olyan kritériumoknak, mint például a geolokációs helymeghatározás kijátszása, hiszen például

[24] WEBSTER 2011.

egy Közel-Kelet ellen irányuló művelet esetén nem célszerű, ha a profilok helymeghatározása mondjuk Coloradóba, a US Air Force Akadémiájához vezet. A VPN mellett az álprofiloknak az adott célterületre testreszabott legendával is kellett rendelkezniük. Ehhez hasonló botnethálózatokat az elmúlt években vélelmezhetően több állam is kiépített. Az e célra létrehozott botnetek különösen a Twitteren népszerűek.[25] A botnethálózatok mellett az egyes államok úgynevezett trollhadseregeket is alkalmaznak. Az egyik legismertebb trollhadsereg Oroszországhoz köthető. Egykori tagok beszámolói alapján ezek a műveletek szigorúan szabályozott keretek között működnek.[26] A Szentpéterváron található Internet Research Agency nevű, online kutatással foglalkozó cégnél például váltott műszakban,[27] hármas csoportokban,[28] eltérő bérkategóriába[29] sorolva dolgoznak becslések szerint ezren, hogy Nyugat-ellenes, Kreml-barát híreket osszanak meg hazai és külföldi portálokon.[30] A témákat az adott nap elején jelölik ki, és meghatározott számú kommentet[31] kell meghatározott számú profillal elhelyezni. Ezeket nagyban meghatározzák az aktuális kül- és belpolitikai történések. Természetesen nem csak Oroszország tart fenn ilyen trollhadseregeket. Kína esetében több millió személyből álló csoportokról beszélhetünk,[32] de vélelmezhetően nyugati államok is építettek ki ehhez hasonló képességet.

A Twitteren 2006 és 2017 között terjedő hírek vizsgálatából Sinan Aral és szerzőtársai arra a megállapításra jutottak,[33] hogy az álhírek gyorsabban, szélesebb körben, távolabbra és mélyebben terjedtek az összes megfigyelt információs kategóriában, egyes esetekben jelentősen nagyobb mértékben előzték meg a hiteles hírek terjedését. Ezenfelül azt is igazolták, hogy a hamis politikai hírek kiemelkednek a többi álhír közül, ezek külön gyorsabban, szélesebb körben, mélyebben terjednek. Vizsgálataik azt is bebizonyították, hogy az emberek gyorsabban terjesztik az álhíreket, mint a botnetek.

Esettanulmány

Donald Trump elnökké választásában számos okot azonosíthatunk, ezeknek csupán egy szeletét jelentik az amerikai nemzetbiztonsági szolgálatok Oroszországot vádoló befolyásoló műveletei, illetve a közösségimédia-felületein terjesztett álhírek. A Facebook 2017 szeptemberében elismerte, hogy 2015 júniusa és 2017 májusa között mintegy 470, feltehetően Oroszországból való, hamis profillal vagy oldallal bejelentkező felhasználó mintegy százezer dollárt költött közel háromezer hirdetés megjelentetésére. A hirdetések nem csupán politikai tartamúak voltak, de például a bevándorlóellenességre építő álhírek, tartalmak,

[25] Abokhodair–Yoo–McDonald 2015.
[26] Walker 2015.
[27] Helyiségenként hozzávetőlegesen 20 fő dolgozott 3 szerkesztő alá sorolva.
[28] Ebből volt egy témafelvető, akihez később csatlakoztak a többiek vitát generálva, megerősítve a hírt stb.
[29] 2015-ben ez 45 ezer rubelnek (219 ezer forintnak) megfelelő havi bérezést, angol nyelvű kommentek esetében 65 ezer rubelt (316 ezer forint) jelentett.
[30] A leggyakoribb visszatérő elem, hogy a Nyugat, az európai civilizáció a vesztébe rohan a dekadencia, a liberalizmus, újabban a menekültek, a gyenge vezetők miatt, az egyetlen mentsvár az erőskezű, feddhetetlen Vlagyimir Putyin.
[31] 12 órás műszakban 135 kommentet.
[32] Yang 2017.
[33] Vosoughi – Roy–Aral 2018.

illetve a társadalmat megosztó kérdésekkel kapcsolatban is számos tartalmat osztottak meg (például fegyvertartás, polgárjogi mozgalmak stb.) A Facebook belső vizsgálata megállapította, hogy a közösségi oldalon 29 millió amerikai felhasználó került közvetlenül interakcióba a vonatkozó tartalmakkal, és összesen 126 millió amerikai felhasználóhoz jutottak így el a hirdetések, bejegyzések.[34] Ezenfelül a Facebook 170 Oroszországhoz köthető Instagram-fiókot törölt, amelyek nagyjából 120 ezer témába vágó bejegyzést tettek közzé.

De nem a Facebook volt az egyetlen, amely fizetett hirdetések formájában segítette az orosz lélektani műveleteket. A Twitter 2752 Oroszországhoz köthető csatornát, illetve további 50 ezer automatikusan tweetelő álprofilt azonosított, amelyek összesen 1,4 millió bejegyzést tettek közzé a kampány során.[35] A Google is érintett volt ez ügyben, a YouTube-ra 1108 darab vonatkozó videót töltöttek föl 43 órát kitevő tartalommal. Ezek terjesztését bizonyíthatóan 4700 dollárnyi hirdetéssel indították be.[36] Ezek a hirdetések azért voltak annyira sikeresek, mert a Cambridge Analytica képes volt az általa kidolgozott módszertannal a lehető legpontosabban célozni a felhasználókat.

Tanulmányunkhoz egy olyan álhír terjedésének vizsgálatát végeztük el, amely egyrészt a lélektani műveletek tervezésével kapcsolatos, a fentebb ismertetett három módszer mindegyikét tartalmazza, másrészt a fizikai térben is komoly következményekkel járt.

2016. október 29-én egy Carmen Katz nevű Facebook-felhasználó megosztott egy hírt a New York-i rendőrségen dolgozó ismerősére hivatkozva, amelynek értelmében a washingtoni Comet Ping Pong pizzériában Hillary Clinton, John Podesta és egy, a Demokrata Párt elitjéhez köthető hálózat sátánista rituálék keretében pedofil tevékenységeket végez. A hír alapjául John Podesta, Clinton kampányfőnökének 2016 márciusában feltört e-mailjei szolgáltak. A Demokrata Párt szervereit és Podesta fiókját ért kibertámadással az amerikai nemzetbiztonsági szolgálatok Oroszországot vádolták,[37] majd októberben a WikiLeaks publikálta Podesta levelezéseit. A WikiLeakset egyre gyakrabban éri az a vád, hogy valójában Oroszország irányítása alatt működik, a nyilvánosságra hozott információk orosz hackertámadások következtében kerülnek Julien Assange birtokába. Bár mind a WikiLeaks, mind Oroszország tagadja ezeket a vádakat, számos jel utal arra, hogy nem alaptalanok az ezzel kapcsolatos vádak. A WikiLeaksre kiszivárogtatott információk közé könnyű olyan információkat is becsempészni, amelyek nem valósak, és egy kampány keretében ilyen valótlanságokat tartalmazó híreket nem nehéz elterjeszteni, hogy az egyébként ellenőrzött, valós információk hitelessége ezeket az álhíreket is validálja. Ez a tevékenység mind a reflexív kontroll módszere, mind a társadalmi vírus kapcsán érvényes.

Katz posztjából indult útnak az úgynevezett #pizzagate botrány.[38] Maga az ügy az összeesküvés-elméletek számos elemét magát viselte. Fontos különbséget tenni az összeesküvés-elméletek és a konspirációs teóriák (konteók) között, ugyanis a köznyelvben gyakran szinonimaként használják ezeket. A különbséget Tóth Tibor, tiboru, a *konteó* szó megalkotója, a konteó blog írója az alábbiak szerint fogalmazta meg: „Az összeesküvés elméletek elvakult hívei azok, akik válogatás és kritika nélkül mindent elhisznek: a gyíkemberré átváltozó angol királyi családtól a Hold túloldalán létező náci bázisokig. Mi, konteósok pedig

[34] Stamos 2017.
[35] Swaine 2018.
[36] Timberg–Dwoskin 2017.
[37] Kovács–Krasznay 2017.
[38] Robb 2017.

minden elméletet egészséges gyanakvással fogadunk és megkérdőjelezünk. Megpróbálunk a színfalak mögé nézni, de nem hisszük, hogy az emberiség sorsa azon áll vagy bukik, hogy a Bilderberg-csoport következő ülésén arról döntenek, hogy meg kell-e hülyíteni az embereket. A konteózás az én elméletem szerint intellektuális kihívás, szellemi játék, és ha megfelelő társaságba kerül az ember, nagyon jó hangulatú beszélgetések alakulnak ki."[39] Mindez azért is fontos, mert a tartalomfogyasztási szokásokban komoly következménnyel jár. Amennyiben valaki kritikus az olvasott hírekkel szemben, különböző, egymásnak ellentmondó forrásokat is elolvas, ez pedig csökkenti a „bubble filter"-jelenség kialakulását.

A #pizzagate egy rendkívül szövevényes összeesküvésről szólt, amelynek terjesztésében a 4 Chan nevű fórum is aktívan részt vett. A 2003-ban létrehozott 4 Chan egy fórumszolgáltatás, ahol az alapvetően anonim felhasználók különböző tartalmakat osztanak meg egymással. Az oldal relevanciája a hacktivizmusban ragadható meg, az Anonymous hackercsoport[40] az oldal felhasználóiból rekrutálódott. A 4 Chan névtelen, azóta már törölt fórumozója szerint Podesta e-mailjeiből dekódolhatók a sátánista, pedofil rituálék. Ez alapján például az e-mailekben szereplő hot dog a fiú, a pizza a lány, a sajt a kislány, a tészta a kisfiú, a jégkrém a fiú prostituált, a dió a színesbőrű egyén, a kolbász az orgia kódszavai. A lelkes fórumozók ezek ismeretében átböngészték a nyilvánosságra hozott e-maileket, és az ezeket a szavakat tartalmazó üzeneteket bizonyítékként könyvelték el a sátánista, pedofil tevékenységre. Hogy ezek mennyire reálisak, az alábbi példa kiválóan szemlélteti: „Hahó, John! Tudjuk, hogy a konyhaművészet mestere vagy, ezzel már hosszú évek óta tisztában vagyunk… de mégis, diószósz tésztára? Mary, könyörgök, légy őszinte, valóban ízletes volt az a szósz?"

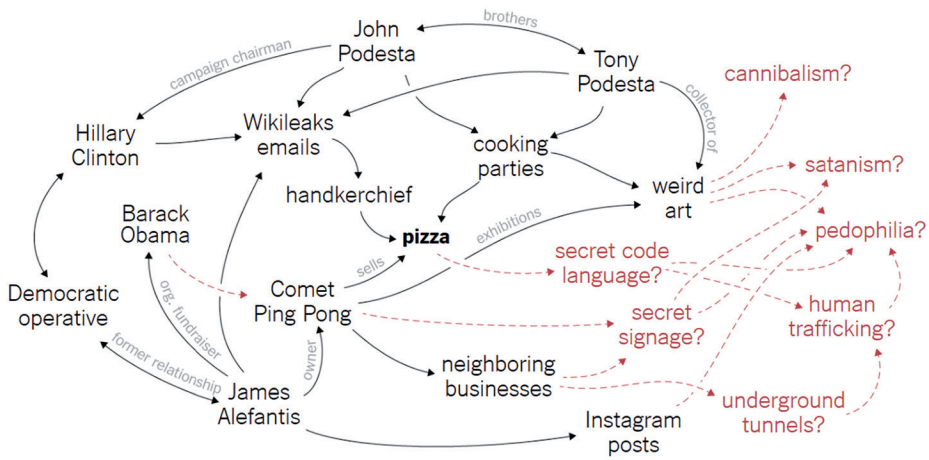
A sátánista irányultságot a pizzéria logója „bizonyította" (lásd a 2. számú ábrát), amelyből egyesek a radikális iszlamista kapcsolatot is felfedezni vélték. A későbbiekben már Barack Obamát is kapcsolatba hozták az üggyel, amely embercsempészettel, kannibalizmussal és egyéb tevékenységekkel egészült ki az interpretációk során. A *The New York Times* mindezt egy látványos hálózatban ábrázolta (lásd a 3. számú ábrát).

[39] Szegedi 2015.

[40] Az Anonymous hacktivista közösség decentralizált, egymáshoz lazán köthető csoportok globális hálózata. Kezdetben az internet cenzúrája ellen, az internet szabadságáért harcoltak, később a fennálló világrend megdöntését, a politikai és gazdasági rendszerek átalakítását tűzték ki céljuknak. Célpontjuk volt többek között a Szcientológia Egyház, a Sony, a Los Zetas mexikói drogkartell, a WikiLeakset bojkottáló pénzügyi vállalatok, az iráni, egyiptomi, tunéziai kormány, újabban az Al-Káida és az Iszlám Állam. Bővebben lásd: Berki 2013.

2. ábra
A Comet Ping Pong Pizzéria bejárata

Forrás: AISCH–HUANG–KANG 2016



3. ábra
A #pizzagate szereplői

Forrás: AISCH–HUANG–KANG 2016

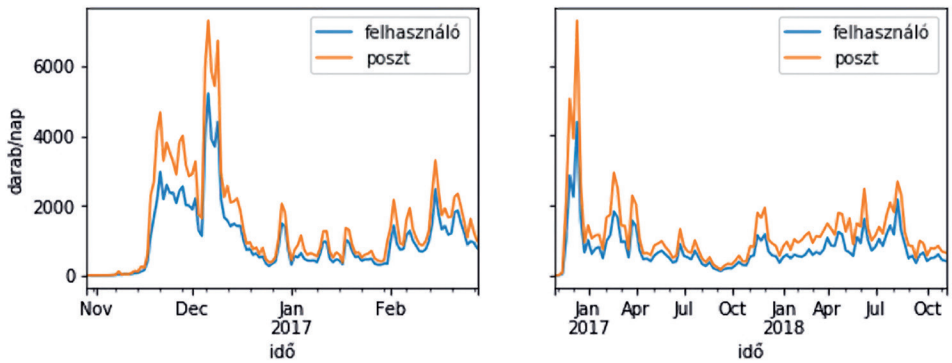
2016 decemberében a YouGov nevű kutatóintézet vizsgálta meg bizonyos összeesküvéselméletek társadalmi elfogadottságát. A vizsgálatnak részese volt a #pizzagate is, ami meglepő eredményt hozott, hiszen nemcsak a Trump-szavazók 46%-a hitte igaznak, hanem a Clinton-szavazók 17%-a is. Ennek veszélyességét Edgar Maddison Welch esete is jól demonstrálta. Welch számos posztot megosztott a #pizzagate-üggyel kapcsolatban, teljes egészében igaznak gondolta a benne szereplő híreket, majd úgy döntött, saját kezébe veszi az igazságszolgáltatást. 2016. december 1-én elutazott Washingtonba, majd gépkarabéllyal lövöldözni kezdett a pizzériában. Az esetben szerencsés módon senki nem sérült meg, Welchet végül négyéves börtönbüntetésre ítélték.[41]

Úgy véljük, a #pizzagate egy meglehetősen komplex álhír, amely több szempontból is érdekes lehet a hálózatelemzés kapcsán. A *#pizzagate* kifejezést tartalmazó megosztások vizsgálatát a *SentiOne* nevű, szentimentanalízist alkalmazó oldal segítségével végeztük. A SentiOne (www.sentione.hu) teljes Európát lefedő, 30 nyelven beszélő és webes szöveganalitikán alapuló social listening szoftvere kulcsszavas keresés alapján, valós időben vagy akár három évre visszamenően figyeli, indexálja és elemzi az internetes fórumokon, blogokon, weboldalakon és közösségimédia-csatornákon közzétett publikus szöveges tartalmak minden típusát, amelyek önmagukban vagy kontextusukban tartalmazzák a felhasználó által már előre definiált és a platformra felvitt kulcskifejezések bármelyikét. A releváns tartalmakat kvantitatív kutatás céljából és ezt megkönnyítendő, különböző fókuszpontok és kutatási paraméterek mentén rendezi össze, amelyeket interaktív grafikonokon ábrázol. A kvalitatív mélyelemzéseket is támogató módszertani, technológiai felépítés pedig biztosítja a kutatáshoz kapcsolódó összes indexált tartalom, poszt, komment, cikk és említés egyenként történő elemzésének és kategorizálásának lehetőségét is. A keresés idősávjának kezdetét 2016. október 29-e, Katz posztjának megjelenése adta, 2018. november 10. volt a záróidőpont. Az erre az idősávra szűkített keresés során a *pizzagate* keresőkifejezést alkalmaztuk. A keresés kontextusérzékeny volt, vagyis azokat a találatokat is listázta, amelyben szó szerint nem szerepelt a *pizzagate* kifejezés, azonban a megosztás ezzel kapcsolatos diskurzust tartalmazott.

Ily módon a SentiOne által indexált oldalak esetében összesen 677 773 tételből álló adatbázist kaptunk. A SentiOne a megosztás esetében vizsgálja azt is, hogy a megosztó milyen érzelem kifejezésével posztolta a tartalmat. Ez alapján 43 669 (6,52%) esetben pozitívan, 100 041 (14,94%) esetben negatívan írtak a #pizzagate kapcsán. A fennmaradó 543 063 esetet semlegesnek jelölte a rendszer, ami nemcsak a neutrális érzelemre utal, hanem sok esetben nem tudta értelmezni az érzelem irányát. A megosztók jelentős többsége férfi volt (61,31%).

A SentiOne oldaláról az online grafikonos elemzés mellett a szöveges adatok is letölthetők további offline elemzésre. Az általunk vizsgált *pizzagate* keresőszóhoz kapcsolódó első találat 2016. október 29-én jelent meg, és az adatgyűjtés időpontjáig aktívnak mutatkozott:

[41] Robb 2017.



4. ábra

A SentiOne által indexelt online médiában a pizzagate *kifejezés heti szinten aggregált megjelenéseinek a száma*

Megjegyzés: Sárga szín jelöli a megjelent bejegyzések (posztok) számát, kék pedig a szerzők számának becsült értékét. Bal oldalon az időtengely ki van nagyítva a legaktívabb időszak környékén.

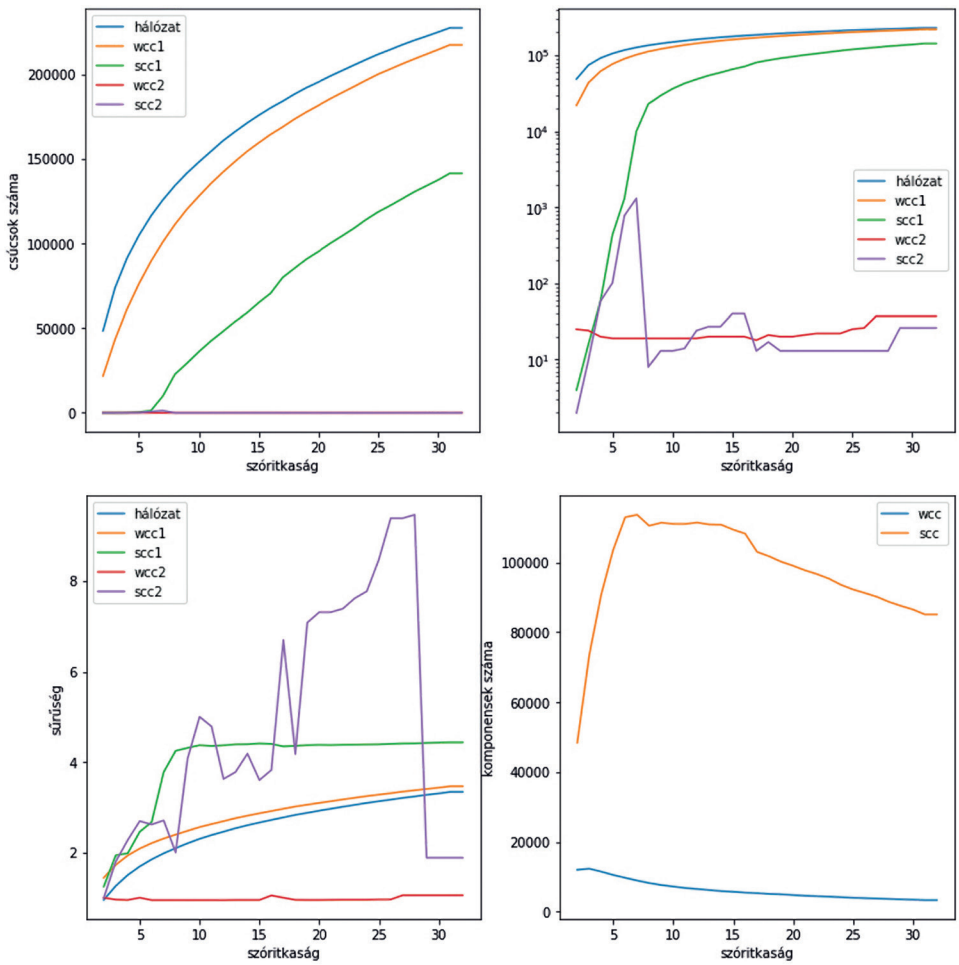
Forrás: a szerző szerkesztése

A következőkben olyan hálózatok tulajdonságait elemezzük, amelyeket a témához kapcsolódó médiabejegyzések alapján definiáltunk, és a felvetett hír terjedését jellemzik.

A poszthálózat

A médiabejegyzések közti kapcsolatot nagyon sokféleképpen lehet definiálni. Amennyiben az egyes bejegyzésekről a médiaszolgáltató által kezelt egyéb azonosítók is rendelkezésre állnak (úgynevezett metaadatok), akkor ezek között többnyire található egy egyedi azonosító is minden egyes szöveghez. Ha valaki idézi az adott bejegyzést, vagy válaszol rá, akkor a médiaportálok a metaadatokban ezt is nyilvántartják. Így megbízhatóan követhető egy-egy hír terjedésének útja. Ha azonban egy-egy hír terjedését a különböző portálok között is követni szeretnénk, akkor ezek a metaadatok már nem használhatók, csupán a hír szövegére vagyunk utalva. Az irodalomból ismert (Menzer) egyik módszer azt használja ki, hogy ha valaki átvesz egy hírt egy korábbi bejegyzésből, akkor jó eséllyel idéz is belőle, vagy legalábbis a szóhasználatában hasonlítani fog az eredeti bejegyzésre. Számos szöveghasonlósági mérték mellett, amelyek kiértékelése akár igen számításigényes is lehet, egy igen egyszerű megközelítés, ha a hálózati kapcsolatokat a közösen használt, de egyébként ritka szavak felbukkanása alapján határozzuk meg. A hálózatunk tehát a médiabejegyzések hálózata lesz, ahol a csúcsok az egyes posztok, két csúcs pedig akkor van összekötve egy irányított éllel, ha a két bejegyzésben van legalább egy közös, de ritkán használt szó. Az élek irányítottságát a bejegyzések megjelenési időpontja határozza meg: a nyíl iránya a korábbitól a később megjelent felé mutat. Amennyiben két poszt időbeli sorrendje nem határozható meg egyértelműen (például egy blogoldalon a megjelenés időpontját csak óra

pontossággal jelzik), akkor a két poszt közé mindkét irányba mutató él került. Az alábbi ábra szemlélteti az így definiált hálózatok méretének néhány tulajdonságát. Látható, hogy amint egyre gyakoribb szavak alapján kötjük össze a csúcsokat, úgy egyre nagyobb hálózatokat kapunk, amelyek egyre sűrűbben vannak összekötve. Érdemes megjegyezni azt is, hogy a hálózatokban a sűrűsödéssel az elkülönülő komponensek száma csökken, és a hálózat egy domináns óriáskomponenst tartalmaz.



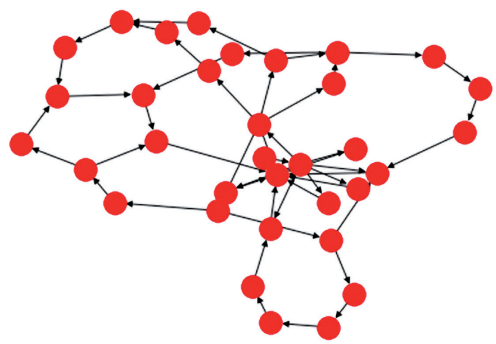
5. ábra

Közösen használt ritka szavakkal definiált hálózatok méretei és sűrűségei a kapcsolatként használt szó előfordulásának (ritkaságának) függvényében

Megjegyzés: A wcc a gyengén összekötött, míg az scc az erősen összekötött komponensre utal. A wcc1 és scc1 a legtöbb csúcsot tartalmazó, míg a wcc2 és scc2 a második legnagyobb komponenst jelöli.

Forrás: a szerző szerkesztése

A kialakult hálózatok struktúráját szemlélteti az alábbi ábra, ahol az egyik erősen összekötött komponenst ábrázoltuk. Érdemes megjegyezni, hogy ilyen nagy, erősen összekötött komponens, azaz ahol vannak hurkok az irányított hálózatban, csak durva időfelbontás esetén alakulhat ki. Amennyiben a posztok megjelenési idejét a másodperc tört részének pontosságával ismernénk (például a Twitter esetén), akkor csak robottweetelők esetén fordulhatna elő, hogy ilyen sok, azonos időben megjelenő bejegyzést látunk.



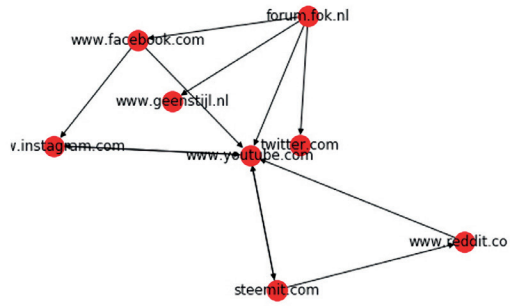
6. ábra

A legnagyobb, erősen összekötött komponens a közösen használt, ritka szavakkal (4 előfordulással rendelkező szavak) definiált internetes bejegyzések hálózatában

Megjegyzés: Érdekes megfigyelni az irányított hurkokat, amelyek egy központi blogbejegyzéshez vezetnek. E bejegyzés időpontja csak nagyon pontatlanul ismert, ezért lehetséges a komponens létezése.

Forrás: a szerző szerkesztése

Egy-egy internetes bejegyzés azonban nagyon egyedi, a felhasználók csak igen kis részéhez jut el közvetlenül. A posztokat összekötő hálózat tehát egy igen részletes felbontású elemzést ad a terjedési eseményekről. Ha nagyobb léptékű áttekintést szeretnénk kapni, akkor a hálózat részeit össze kell vonnunk, nagyobb egységeket kell csúcsokként definiálni. Internetes írások esetén az egyik aggregálási lehetőség, ha azt vizsgáljuk, hogy az egyes bejegyzések melyik portálokon jelentek meg. Az alábbi ábra a *pizzagate* kulcsszóval előforduló írásokat közzétevő domainnevek hálózatát mutatja. Itt az egyes csúcsok azok az internetes webcímek, ahol az egyes írások megjelentek. A csúcsok közti kapcsolatokat a korábban definiált bejegyzések közti hálózat alapján vezetjük le: ha két portál megjelentetett legalább egy-egy olyan cikket, amelyek össze vannak kötve a közös ritka szó elve alapján, akkor a portálokat reprezentáló csúcsokat a cikkeket összekötő él irányítottságának megfelelően összekötjük egymással.

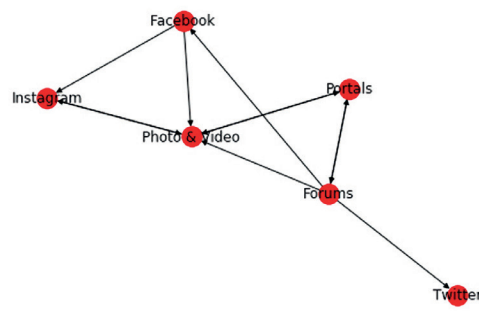


7. ábra

A második legnagyobb, gyengén összekötött komponens portálszintű kapcsolati hálózata a 2 előfordulású szavakkal definiált hálózatban

Forrás: a szerző szerkesztése

Még magasabb szintű áttekintést nyerhetünk, ha az egyes internetes médiumokat típus szerint összevonjuk. Az alábbi ábra egy ilyen, magasabb rendű összevonással kapott hálózatot mutat. Érdekes megfigyelni, hogy még ilyen magas szintű összevonás után is maradnak egyirányú élek, azaz vannak olyan portálok, ahol a ritkább szavakat használó bejegyzések később jelennek meg. Feltehetően ez az áramlási irányultság az egyes médiatípusok eltérő stílusával, használati módjával függhet össze.



8. ábra

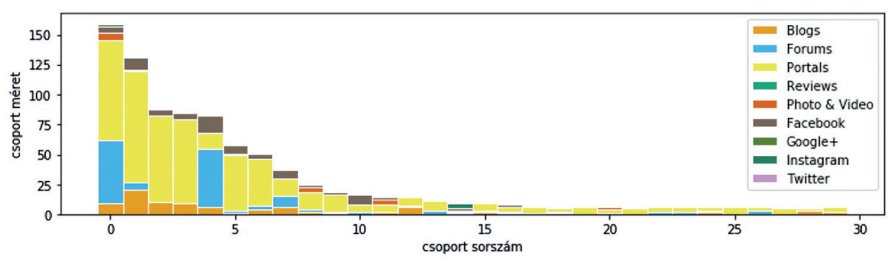
Az internetes hírek terjedésének magas szintű, áttekintő hálózata

Megjegyzés: Médiatípusok közötti áramlás a második legnagyobb, gyengén összekötött komponens alapján a 2 előfordulással rendelkező közös szavakkal definiált hálózatban.

Forrás: a szerző szerkesztése

A különböző médiatípusok eltérő szerepe más módon is megfigyelhető. Egy hálózatnak vannak sűrűn és kevésbé sűrűn összekötött részei, azaz vannak olyan csúcsok, amelyek között több él fut, és vannak, amelyek kevesebb éllel vannak összekötve. A hálózatnak azokat a részeit, amelyek erősebben össze vannak kapcsolódva, úgynevezett csoportkereső eljárásokkal lehet megkeresni. Ezek közül az egyik lehetséges algoritmus a hálózat legsűrűbben összekötött csúcsait keresi. A legsűrűbb akkor lehet egy hálózat, ha minden csúcsa minden másik csúcsával össze van kötve. Kis (5-10 csúcsból álló) hálózatok esetén ez még a gyakorlatban is sokszor előfordul, nagyobb hálózatok esetén azonban csak ritka kivétel, ha minden mindennel kapcsolatban áll. Ezért enyhíthetünk a sűrűség feltételén úgy, hogy megkeressük a hálózatban azokat a részeket, amelyekben mindenki mindenkivel összekötött, és azt tekintjük csoportnak, amit ezekből a maximálisan összekötött, egymással átfedésben lévő részekből állítunk össze. Ha csak a hálózat sűrűségét vesszük feltételnek a csoportok kialakításához, akkor a hálózatok éleinek irányától el szoktunk tekinteni, azaz a megfelelő irányítatlan hálózat szerint csoportosítjuk a csúcsokat.

Az internetes posztok két előfordulással rendelkező szavak átfedésével definiált hálózatának legnagyobb, gyengén összekötött komponensében kerestünk sűrű csoportosulásokat. A megtalált csoportokban megvizsgáltuk, hogy milyen médiatípusban jelentek meg a csoportot alkotó csúcsok. Az alábbi ábra hisztogramja a csoportok méretét és ezen belül az egyes médiatípusok előfordulási arányát mutatja. Látható, hogy a csoportok nagy része fórumok és portálok szövegeiből áll össze. Ez a tulajdonság az egyes médiatípusokban megjelenő átlagos szöveg hosszával függhet össze. A hálózatunk definiálásakor ugyanis olyannyira ritka szavakat használtunk, hogy egy szó csak egyetlen élt hozhatott létre a hálózatban. Több csúcs kapcsolódásához arra volt szükség, hogy egy-egy szövegben, amelyet egy hálózati csúccsal reprezentáltunk, több ritka szó is előforduljon. Azon médiában, ahol gyorsan és gyakran publikálnak a felhasználók, kis eséllyel fordulnak elő a ritka szavak. Illetve, ha előfordulnak, akkor jó eséllyel valamilyen korábbi hírben olvasott szó átvételéről van szó (lásd még a 6. ábrát).



9. ábra

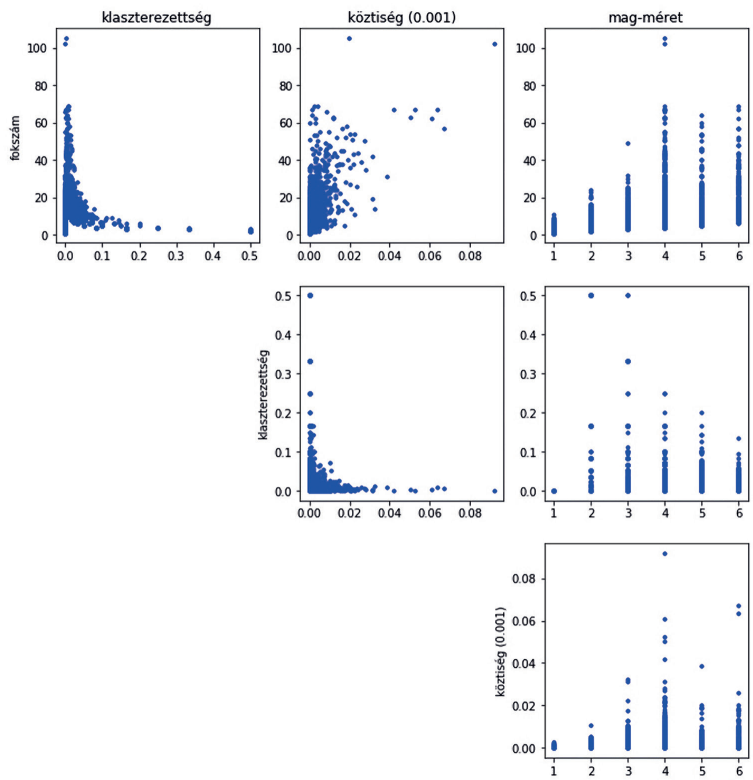
Az interneten megjelenő írások hálózatának csoportméretei és a csoportokon belüli médiatípus-arányok

Megjegyzés: A hálózatot két előfordulással rendelkező ritka szavak határozzák meg.

Forrás: a szerző szerkesztése

Elemzésünket egy korrelációs diagrammal zárjuk, amely a különböző hálózati jellemzők közti kapcsolatot szemlélteti az internetes médiamegjelenések hálózatában. Az elemzés során az irányítottságtól eltekintettünk. A legfelső sorban a fokszámhoz való viszonyt látjuk. A klaszterezettség alacsony értéket vesz fel magas fokszám esetén, és fordítva:

magas klaszterezettség csak alacsony fokszám esetén fordul elő. Ez egy természetes jelenség, hiszen a klaszterezettség akkor magas, ha a csúcsból kiinduló élek végpontjain lévő csúcsok szintén össze vannak kötve. Egy ritka hálózatban kicsi az esélye, hogy sok élpárnak a végpontjai között is legyen egy él. A klaszterezettség és a köztiség összefüggése már nem ilyen nyilvánvaló. Az a megfigyelés, hogy a magas köztiségű csúcsok alacsony klaszterezettséggel járnak, arra utal, hogy a hálózatban az egyes részek közti kommunikáció nem egy sűrűn összekötött magon keresztül folyik, hanem egy vékony hídon vagy egy magas fokszámú csomóponton. A fokszámköztiség-grafikon alapján pedig azt látjuk, hogy az utóbbi, azaz a magas fokszámú csúcs jelenti a fontos összeköttetést. Ha mindehhez a magméret-grafikonból látható összefüggéseket is hozzávesszük, megállapíthatjuk, hogy az e hálózaton terjedő híreket viszonylag egyszerű volt befolyásolni. Ugyanis a hálózaton átmenő legtöbb útvonal által érintett csúcsból kevés van, ráadásul ezek a csúcsok a lokális környezetükben is nagyon aktívak (sok kapcsolattal rendelkeznek). A hírterjesztés gondos tervezettségének pedig az lehet a jele, hogy a fontos csúcsok (magas fokszám és köztiség) körbe vannak véve tartalékszerepű csúcsokkal. Ugyanis ezek a fontos szerepű csúcsok részei egy hármas magszámmal rendelkező alhálózatnak, azaz ebből az alhálózatból bármely csúcs kiszakításához legalább három különböző élt kell megsemmisíteni.



10. ábra

Hálózatos alapmennyiségek kapcsolata a két előfordulással rendelkező szavak által definiált hálóban

Forrás: a szerző szerkesztése

Összefoglalás

Tanulmányunkban a #pizzagate néven ismertté vált összeesküvés-elméletet vizsgáltuk a hálózatok aspektusából. A közösségi oldalak és az álhírek politikai döntéshozatalban betöltött szerepére a 2016-os amerikai elnökválasztás hívta fel a szélesebb közvélemény figyelmét. Egy-egy álhír megfelelő időben történő terjesztése rendkívül hatékonyan képes befolyásolni a közvéleményt. Ehhez persze ismerni kell az álhírek terjedésének mintázatait, azokat a hálózatokat, amelyeken keresztül a leghatékonyabban lehet terjeszteni őket. Az álhírek azért is hatékonyak, mert rendkívül nehéz védekezni velük szemben. Napjainkra sem találták meg azt a módszert, amelynek segítségével hatékonyan lehetne csökkenteni hatásukat. Ahogy a közösségi oldalak valamit változtatnak az algoritmusaikon, azok hatása nagyobb károkat okozhat, mint amire elsőre számítottak. A bevezetőben megfogalmazott digitális ökoszisztéma itt is tetten érhető. Amikor a Facebook rákényszerítette a médiacégeket, hogy elsődlegesen a Facebookon keresztül jelenjenek meg reklámrészesedésért cserébe, a vállalatok ehhez alkalmazkodva stratégiát váltottak, és olyan jellegű tartalmak gyártásába kezdtek, amelyek inkább a közösségi médián való megjelenést támogatják. 2018-ban a Facebook bejelentette az álhírek elleni harc jegyében, hogy a hírfolyamon az ismerőseink életével kapcsolatos tartalmakat fogják preferálni, és az algoritmus alapvetően ezeket a tartalmakat jeleníti meg, háttérbe szorítva a híroldalakat. Ez a lépés azonnal számos kritikát hozott, hiszen ily módon nemcsak az álhíreket terjesztő oldalak elérhetősége csökkent radikálisan, hanem azoké az oldalaké is, amelyek nem fizetnek a megjelenésért. A vádak szerint ez komoly demokratikus deficithez vezet, hiszen ha egy híroldalnak nincs olyan pénzügyi háttere, amivel a megjelenésért fizetni tudna, akkor az általa publikált tartalmak megjelenése veszélybe kerülhet. A demokrácia leépülését vizionálják az egyes kormányzati jogalkotásokkal szemben is. Németországban 2017 nyarán egy olyan jogszabályt fogadtak el, amely alapján maximum 50 millió eurós büntetést szabhatnak ki a közösségi oldalakra, ha a bejelentést követő 24 órán belül[42] nem távolítják el a gyűlöletkeltésre alkalmas tartalmakat. A jogszabály 2018. január 1-jétől vált hatályossá, és minden olyan közösségi oldal esetében kötelező alkalmazni, amelynek legalább 2 millió német felhasználója van. Amennyiben a felhasználó németországi IP-címről keresi fel ezeket az oldalakat, annak látnia kell egy olyan felületet, amin bejelentést tehet, ha gyűlöletkeltésre alkalmas, a német alkotmányt sértő vagy bűncselekményre buzdító posztot lát. Összesen húsz német jogszabály alapján nyílik mód egy bejegyzés jelentésére, beleértve az önkényuralmi jelképek tiltásáról szóló jogszabályt és az alkotmányos rend felforgatásának kísérletét egyaránt. Annak érdekében, hogy a közösségi oldalak eleget tudjanak tenni a törvényi rendelkezésnek, bővítették a moderátorok számát, akiknek el kell dönteni, hogy a jelentett tartalom valóban jogsértő-e, és amennyiben igen, törölniük kell. A jogszabállyal kapcsolatban számos kritikát fogalmaztak meg a német pártok, illetve jogvédők. A legjelentősebb érv az ítélkezés privatizálása, hiszen annak megállapítása, hogy valami törvénytelen-e, vagy sem, megfelelő eljárás keretében a bíróságok feladata. Ezt a feladatot vállalatok nem vehetik át. Ehhez kapcsolódik, hogy rendkívül szoros határidőt szab a döntés meghozatalára, így nincs garancia arra vonatkozóan, hogy az esetlegesen nagy számban jelentett tartalmakat nem törlik-e szinte automatikusan a büntetés elkerülése érdekében, így pedig indokolatlan cenzúra valósulhat

[42] Nem egyértelműen megállapítható tartalmak esetén egy héten belül.

meg. Franciaországban egy 2018 őszén elfogadott törvény alapján az országos választások előtt három hónappal korábban lehetőség nyílik az álhíroldalak cenzúrázására, ami az ellenzék vádja szerint sérti a véleménynyilvánítási szabadságot, illetve a cenzúra esetleges politikai érdekek mentén történő használatára is alkalmas lehet.

Felhasznált irodalom

Abokhodair, N. – Yoo, D. – McDonald, D. W. (2015): Dissecting a Social Botnet. Growth, Content and Influence in Twitter. In *CSCW 2015 – Proceedings of the 2015 ACM International Conference on Computer-Supported Cooperative Work and Social Computing.* New York, ACM. 839–851. DOI: https://doi.org/10.1145/2675133.2675208

Aisch, G. – Huang, J. – Kang, C. (2016): Dissecting the #PizzaGate Conspiracy Theories. *The New York Times,* 2016. 12. 10. Elérhető: www.nytimes.com/interactive/2016/12/10/business/media/pizzagate.html (A letöltés dátuma: 2019. 03. 12.)

AJP-3.7. NATO Military Policy on Psychological Operations (2003). Elérhető: https://info.publicintelligence.net/NATO-PSYOPS-Policy-2003.pdf (A letöltés dátuma: 2019. 03. 12.)

Ált/57, Információs műveletek doktrína. MH DOFT kód: MD 3.10 (1). 2014. 08. 05. 1–15.

Azucar, D. – Marengo, D. – Settanni, M. (2018): Predicting the Big 5 personality traits from digital footprints on social media. A meta-analysis. *Personality and Individual Differences,* Vol. 124. 150–159. DOI: https://doi.org/10.1016/j.paid.2017.12.018

Berki G. (2013): A kibertéri konfliktusok változása. *Hadmérnök,* 8. évf. 1. sz. 173–185.

Briant, E. L. (2018): Pentagon Ju-Jitsu. Reshaping the Field of Propaganda. *Critical Sociology,* Vol. 45, No. 3. 361–378. DOI: https://doi.org/10.1177/0896920517750741

Dudatyev, A. V. (2017): Complex Method of Informational-Psychological Operations Counteraction. *Journal of Automation and Information Sciences,* Vol. 49, No. 1. 76–83. DOI: https://doi.org/10.1615/JAutomatInfScien.v49.i1.70

Eşkisu, M. – Hoşoğlu, R. – Rasmussen, K. (2017): An investigation of the relationship between Facebook usage, Big Five, self-esteem and narcissism. *Computers in Human Behavior,* Vol. 69. 294–301. DOI: https://doi.org/10.1016/j.chb.2016.12.036

Európai Bizottság (2018): Európa digitális fejlődéséről szóló jelentés (EDPR), 2018. Országprofil Magyarországról. Elérhető: http://ec.europa.eu/information_society/newsroom/image/document/2018-20/hu-desi_2018-country-profile-lang_4AA43283-EC48-996F-09918493E34A691F_52334.pdf (A letöltés dátuma: 2019. 03. 12.)

Figueira, Á. – Oliveira, L. (2017): The current state of fake news. Challenges and opportunities. *Procedia Computer Science,* Vol. 121. 817–825. DOI: https://doi.org/10.1016/j.procs.2017.11.106

Haig Zs. – Kovács L. – Ványa L. – Vass S. – Németh A. (2014): *Elektronikai hadviselés.* Budapest, Nemzeti Közszolgálati Egyetem.

Haig Zs. – Várhegyi I. (2005): *Hadviselés az információs hadszíntéren.* Budapest, Zrínyi Kiadó.

He, Q. – Glas, C. A. W. – Kosinski, M. – Stillwell, D. J. – Veldkamp, B. P. (2014): Predicting self-monitoring skills using textual posts on Facebook. *Computers in Human Behavior,* Vol. 33. 69–78. DOI: https://doi.org/10.1016/j.chb.2013.12.026

HOFFMAN, F. G. (2007): *Conflict in the 21st Century. The Rise of Hybrid Wars.* Arlington, Potomac Institute for Policy Studies. Elérhető: www.potomacinstitute.org/images/stories/publications/potomac_hybridwar_0108.pdf (A letöltés dátuma: 2019. 03. 12.)

JANG, S. M. – KIM, J. K. (2018): Third person effects of fake news. Fake news regulation and media literacy interventions. *Computers in Human Behavior,* Vol. 80. 295–302. DOI: https://doi.org/10.1016/j.chb.2017.11.034

KOVÁCS L. – KRASZNAY Cs. (2017): „Mert övék a hatalom". Az internet politikát (is) befolyásoló hatása a 2016-os amerikai elnökválasztás során. *Nemzet és Biztonság,* Vol. 10, No. 3. 3–15. Elérhető: www.nemzetesbiztonsag.hu/cikkek/nb_2017_3_02_kovacs_laszlo-krasznay_csaba_-_mert_ovek_a_hatalom_-_az_internet_politikat_is_befolyasolo_hatasa_a_2016_os_amerikai_elnokvalasztas_soran.pdf (A letöltés dátuma: 2019. 03. 12.)

LEWANDOWSKY, S. – ECKER, U. K. H. – COOK, John (2017): Beyond Misinformation. Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition,* Vol. 6, No. 4. 353–369. DOI: https://doi.org/10.1016/j.jarmac.2017.07.008

MILLER, D. (2015): Sociology, Propaganda and Psychological Operations. In SMITH, A. – DAWSON, M. – FOWLER, B. – MILLER, D. – SMITH, A. eds.: *Stretching the Sociological Imagination. Essays in Honour of John Eldridge.* London, Palgrave Macmillan. 163–188. DOI: https://doi.org/10.1057/9781137493644_9

NORMAN, W. T. (1963): Toward an adequate taxonomy of personality attributes. Replicated factor structure in peer nomination personality ratings. *The Journal of Abnormal and Social Psychology,* Vol. 66, No. 6. 574–583. DOI: http://dx.doi.org/10.1037/h0040291

PAUL, C. – MATTHEWS, M. (2016): *The Russian „Firehose of Falsehood" Propaganda Model. Why It Might Work and Options to Counter It.* RAND Corporation. Elérhető: www.rand.org/content/dam/rand/pubs/perspectives/PE100/PE198/RAND_PE198.pdf (A letöltés dátuma: 2019. 03. 12.)

PEGLAR, T. (2018): 1995 Reintroduction of Wolves in Yellowstone. *YellowStonePark.com,* 2018. 07. 09. Elérhető: www.yellowstonepark.com/park/yellowstone-wolves-reintroduction (A letöltés dátuma: 2019. 03. 12.)

PIX G. (2005): *A lélektani műveletek jellemzőinek vizsgálata.* Doktori értekezés kézirata, Budapest, Zrínyi Miklós Nemzetvédelmi Egyetem.

ROBB, A. (2017): Anatomy of a Fake News Scandal. *Rolling Stone,* 2017. 11. 26. Elérhető: www.rollingstone.com/politics/politics-news/anatomy-of-a-fake-news-scandal-125877/ (A letöltés dátuma: 2019. 03. 12.)

ROCHE, E. M. (2017): Comments on "Assessing Russian Activities and Intentions in Recent US Elections". *Cyberarmscontrolblog,* 2017. 01. 08. Elérhető: https://cyberarmscontrolblog.com/2017/01/08/comments-on-assessing-russian-activities-and-intentions-in-recent-us-elections/ (A letöltés dátuma: 2019. 03. 12.)

STAMOS, A. (2017): An Update On Information Operations On Facebook. *Facebook Newsroom,* 2017. 09. 06. Elérhető: https://newsroom.fb.com/news/2017/09/information-operations-update/ (A letöltés dátuma: 2019. 03. 12.)

SWAINE, J. (2018): Twitter admits far more Russian bots posted on election than it had disclosed. *The Guardian,* 01. 20. Elérhető: www.theguardian.com/technology/2018/jan/19/twitter-admits-far-more-russian-bots-posted-on-election-than-it-had-disclosed (A letöltés dátuma: 2019. 03. 12.)

Symeonidis, I. – Shirazi, F. – Biczók, G. – Pérez-Solà, C. – Preneel, B. (2018): Collateral damage of Facebook third-party applications. A comprehensive study. *Computers & Security,* Vol. 77. 179–208. DOI: https://doi.org/10.1016/j.cose.2018.03.015

Szegedi É. (2015): Mi, magyarok imádjuk az összeesküvés-elméleteket. *SzeretlekMagyarorszag.hu,* 2015. 10. 18. Elérhető: www.szeretlekmagyarorszag.hu/mi-magyarok-imadjuk-az-osszeeskuves-elmeleteket/ (A letöltés dátuma: 2019. 03. 12.)

The Cambridge Analytica Files (é. n.). *The Guardian.* Elérhető: www.theguardian.com/news/series/cambridge-analytica-files (A letöltés dátuma: 2019. 03. 12.)

Timberg, C. – Dwoskin, E. (2017): Russian content on Facebook, Google and Twitter reached far more users than companies first disclosed, congressional testimony says. *The Washington Post,* 2017. 10. 30. Elérhető: www.washingtonpost.com/business/technology/2017/10/30/4509587e-bd84-11e7-97d9-bdab5a0ab381_story.html?noredirect (A letöltés dátuma: 2019. 03. 12.)

Traylor, T. – Freese, C. – Wong, W. (2017): PSYOP, Deception, and Cyberspace in the Open. Analysing Fake News in a Cyber New Normal Communications Environment. In Scanlon, M. – Le-Khac, N.-A. eds.: *16th European Conference on Cyber Warfare and Security (ECCWS 2017).* UK, Academic Conferences and Publishing International Ltd. 488–496.

Vosoughi, S. – Roy, D. – Aral, S. (2018): The Spread of True and False News Online. *Science,* Vol. 359, No. 6380. 1146–1151. DOI: https://doi.org/10.1126/science.aap9559

Walker, S. (2015): Salutin' Putin. Inside a Russian Troll House. *The Guardian,* 2015. 04. 02. Elérhető: www.theguardian.com/world/2015/apr/02/putin-kremlin-inside-russian-troll-house (A letöltés dátuma: 2019. 03. 12.)

Webster, S. C. (2011): Revealed. Air Force ordered software to manage army of fake virtual people. *The Raw Story,* 2011. 02. 18. Elérhető: www.rawstory.com/rs/2011/02/18/revealed-air-force-ordered-software-to-manage-army-of-fake-virtual-people/ (A letöltés dátuma: 2019. 03. 12.)

Yang, Y. (2017): China's Communist party raises army of nationalist trolls. *Financial Times,* 2017. 12. 30. Elérhető: https://www.ft.com/content/9ef9f592-e2bd-11e7-97e2-916d4fbac0da (A letöltés dátuma: 2019. 03. 12.)

Zimdars, M. (2016): *False, Misleading, Clickbait-y, and Satirical "News" Sources.* Elérhető: https://21stcenturywire.com/wp-content/uploads/2017/02/2017-DR-ZIMDARS-False-Misleading-Clickbait-y-and-Satirical-%E2%80%9CNews%E2%80%9D-Sources-Google-Docs.pdf (A letöltés dátuma: 2019. 03. 12.)

Vákát oldal

Krasznay Csaba – Dobos László – Palla Gergely – Pollner Péter

Információbiztonsági incidensek a közigazgatásban

Bevezetés

Az információbiztonsági folyamat egyik legfontosabb részterülete az incidensmenedzsment, azaz azon biztonsági események kezelése, amelyek minden előzetes erőfeszítés ellenére bekövetkeztek, és a szervezet által kívánatosnak tartott biztonsági egyensúlyt megbontják. Ez napjainkban azért is fontos, mert mind a hazai jogszabályok, mind az Európai Unióban elfogadott, így Magyarországra és a magyar közigazgatásra is érvényes kiberbiztonsági szabályozások hangsúlyos elemként tekintenek a bekövetkezett biztonsági események hatóságok felé történő bejelentésére. Ez a kötelezettség pedig csak a megfelelően felépített és működtetett incidensmenedzsment útján teljesíthető.

Az incidensekkel kapcsolatos jogi szabályozás

A nemzetközi szakirodalomban a biztonsági események és a biztonsági incidensek két különböző fogalmat takarnak, azonban a magyar jogszabályi környezetben ezek összemosódnak. A 2013. évi L. törvény az állami és önkormányzati szervek elektronikus információbiztonságáról (a továbbiakban: Ibtv.) 1. § (1) bekezdésében a következőképp határozza meg ezeket a kifejezéseket:

- *Biztonsági esemény:* „nem kívánt vagy nem várt egyedi esemény vagy eseménysorozat, amely az elektronikus információs rendszerben kedvezőtlen változást vagy egy előzőleg ismeretlen helyzetet idéz elő, és amelynek hatására az elektronikus információs rendszer által hordozott információ bizalmassága, sértetlensége, hitelessége, funkcionalitása vagy rendelkezésre állása elvész, illetve megsérül;
- *Biztonsági esemény kezelése:* az elektronikus információs rendszerben bekövetkezett biztonsági esemény dokumentálása, következményeinek felszámolása, a bekövetkezés okainak és felelőseinek megállapítása, és a hasonló biztonsági események jövőbeni előfordulásának megakadályozása érdekében végzett tervszerű tevékenység; […]
- *Súlyos biztonsági esemény:* olyan informatikai esemény, amely bekövetkezése esetén az állami működés szempontjából kritikus adat bizalmassága, sértetlensége vagy rendelkezésre állása sérülhet, emberi életek kerülhetnek közvetlen veszélybe, személyi sérülések nagy számban következhetnek be, súlyos bizalomvesztés következhet be az állammal vagy az érintett szervezettel szemben, alapvető emberi, vagy a társadalom működése szempontjából kiemelt jogok sérülhetnek.”

Az angolszász terminológia a következőképp alakul:[1]

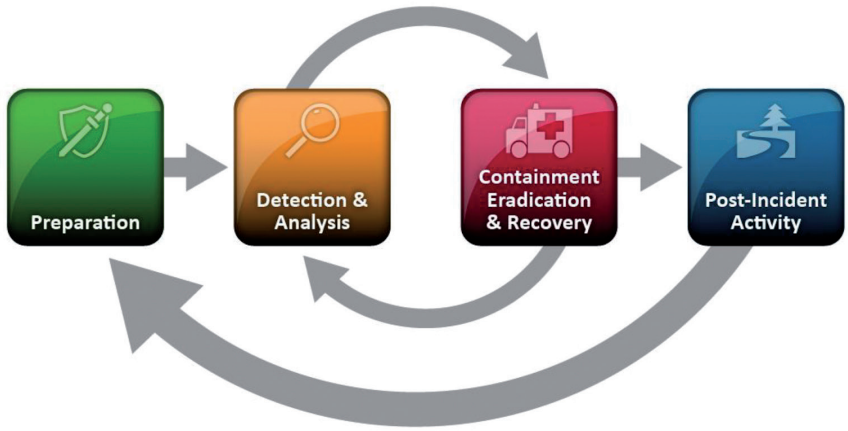
- *Biztonsági esemény:* Minden megfigyelhető előfordulás egy hálózatban vagy egy rendszerben.
- *Biztonsági incidens:* A számítógépes biztonsági szabályzatok, az elfogadható felhasználási irányelvek megsértésének vagy közvetlen fenyegetésének veszélye.
- *Eseménykezelés:* A biztonsági irányelvek és ajánlott gyakorlatok megsértésének mérséklése.

A magyar jogszabály tehát biztonsági esemény alatt valójában a biztonsági incidenseket érti, nem határozza meg az eredeti értelemben vett biztonsági események fogalmát. Ahogy a későbbiekben látni fogjuk, nem minden biztonsági eseményből lesz biztonsági incidens. Egy átlagos magyar közigazgatási szervezetnél a biztonsági események száma naponta több százezer is lehet, ezek túlnyomó többségükben semmilyen problémát nem okoznak, míg a valós incidensek száma nem haladja meg a heti néhány darabot. Természetesen minél felkészültebb egy szervezet, annál nagyobb lesz az incidensek száma, köszönhetően annak, hogy több incidenst képes észlelni és kezelni, azaz több eseményt tud incidenssé minősíteni.

Az incidensmenedzsment tehát egy folyamat, amely egymástól jól elhatárolható lépésekből tevődik össze. Az Egyesült Államok szövetségi információs rendszereinek incidensmenedzsmentjéhez készült NIST SP 800-61 ajánlás a következőképp foglalja össze ezt a folyamatot:

- *Felkészülés:* az incidensmenedzsment szabályozásának, felelősségi körének, eszközrendszerének kialakítása, a különböző szintű (technikai, taktikai, stratégiai) ismeretek beépítése a napi működésbe.
- *Észlelés és elemzés:* az incidensmenedzsment folyamat napi rutinja, amelynek során a biztonsági eseményeket a szervezet észleli és bizonyos esetekben követi, amikor ezek potenciálisan a biztonságot sértő incidenssé válnak, és elemzi, hogy valóban incidensről van-e szó.
- *Elhatárolás, felszámolás és visszaállítás:* az azonosított incidens hatásainak csökkentése, megszüntetése és az eredeti munkafolyamat helyreállítása.
- *Incidens utáni tevékenységek:* Mivel az incidens bekövetkezett, a szervezet védelmi rendszere nem volt képes azt megelőzni, tehát olyan körülmény léphetett fel, amely rámutatott egy biztonsági résre. Ahhoz, hogy ez a rés megszüntethető legyen, elemezni kell az incidenst, és a tanulságok leszűrése után javítani kell a biztonsági rendszert, finomhangolni a beállításokat, oktatni a felhasználókat stb.

[1] Cichonski et al. 2012.



1. ábra
Az incidensmenedzsment folyamata

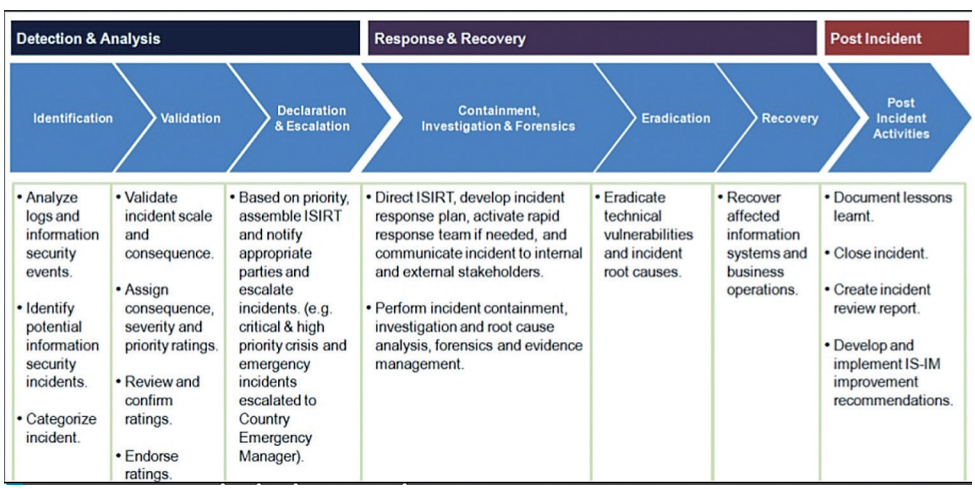
Forrás: Cichonski et al. 2012.

Részletesebben megvizsgálva az ajánlásban leírt folyamatot, kiderül, hogy az ideális incidenskezelés elérése jelentős erőforrások meglétét feltételezi mind humán, mind technológiai oldalról. Ugyanis az egyes folyamatlépéseket tovább bontva olyan feladatok jelennek meg, amelyek több ember specifikus szaktudását feltételezik, felvértezve bizonyos speciális eszközökkel:

- *Észlelés és elemzés:*
  - *Azonosítás:* A biztonsági eseményekről részletes információkat nyújtó informatikai naplóbejegyzések, azaz logok folyamatos elemzése, illetve nagyobb szervezetek esetében a különböző biztonsági eszközökből származó adatfolyamok feldolgozása (akár automatikus, gépi tanulással támogatott megoldásokkal), amely feltételezi az ezen információk gyűjtésére szolgáló infrastruktúra meglétét, valamint olyan személyi állományt, amely képes az adatokból származó információk elemzésére, a potenciális incidensek azonosítására és ezek kategorizálására.
  - *Validáció:* Az azonosított és a kategorizált incidenseket az előzetesen meghatározott szempontok alapján be kell sorolni súlyosság szerint. Először is, meg kell róla győződni, hogy valóban incidensről van-e szó, azaz a biztonsági esemény teljesíti-e a szervezet által előzetesen meghatározott incidenskritériumokat. Amennyiben igen, akkor ellenőrizni kell valószínűsíthető hatásukat, kritikusságukat és prioritásukat, amiben támogatást jelenthetnek a szervezet úgynevezett GRC-eszközei (GRC: governance, risk management, and compliance, azaz irányítás, kockázatmenedzsment és megfelelőség). Ezt a megfelelő felelősségi jogkörrel rendelkező személynek jóvá kell hagynia a részletes információk birtokában.
  - *Tényállás bejelentése és eszkalálása:* a priorizálástól függően megtörténik a szervezeten belül az elhárításhoz szükséges erőforrások értesítése, vagy

kritikus esetben, a jogszabály szerinti súlyos biztonsági esemény esetén a szervezeten kívüli érdekelt feleket is értesítik. Ez lehet például a Kormányzati Eseménykezelő Központ, a Nemzeti Adatvédelmi és Információszabadság Hatóság vagy a rendőrség is.

- *Elhatárolás, felszámolás és visszaállítás:*
  - *Elhatárolás, nyomozás és nyomrögzítés:* Ebben a fázisban a szervezeten belüli erőforrások összehangolt tevékenysége történik. Az információbiztonsági csapaton kívül jellemzően az informatikai üzemeltetés és az érintett folyamatgazdák vesznek részt az elhárításban. Együttesen dolgozzák ki a szükséges lépéseket, miközben gondoskodnak arról, hogy az incidens elhárítása után rendelkezésre álljon minden olyan információ, amelyből a tanulságokat le lehet szűrni, illetve a nyomozó hatóság is tud dolgozni. A cél az incidens hatásainak csökkentése, tehát a továbbterjedés megakadályozása és az érintett eszközök, hálózati szegmensek leválasztása, valamint az incidenst kiváltó ok pontos megértése.
  - *Felszámolás:* Az incidens, az azt kiváltó konkrét sebezhetőség és a gyökérokok megszüntetése sorolható ide.
  - *Visszaállítás:* Az érintett rendszerek és folyamatok visszaállítása az üzleti/közigazgatási terület által is elfogadott eredeti állapotra. Jellemzően napokkal vagy hetekkel az incidens bekövetkezése után lehet ebbe a fázisba eljutni.
- *Incidens utáni tevékenységek:*
  - Ennek során történik meg a tanulságok dokumentálása, az incidens formális lezárása, az incidens formális leírása, valamint azon javaslatok kidolgozása, amelyek a rendszer javítását célozzák, a további, hasonló incidensek elkerülésének érdekében.



2. ábra

Az incidenskezelés részletes feladatlistája

Forrás: AASH 2017

Az Ibtv. és az ahhoz kapcsolódó 41/2015. (VII. 15.) BM rendelet az állami és önkormányzati szervek elektronikus információbiztonságáról szóló 2013. évi L. törvényben meghatározott technológiai biztonsági, valamint a biztonságos információs eszközökre, termékekre, továbbá a biztonsági osztályba és biztonsági szintbe sorolásra vonatkozó követelményekről (a továbbiakban: rendelet) ennek szellemében részletes elvárást támaszt a jogszabály hatálya alá tartozó szervezetek biztonsági eseménykezelésével kapcsolatban. Az Ibtv. már az alapvetéseknél utal arra, hogy a biztonsági események kezelése kiemelt fontosságú tevékenység. Az alapvető elektronikus információbiztonsági követelményekről szóló részben a következőket olvashatjuk:

„5. § Az e törvény hatálya alá tartozó elektronikus információs rendszerek teljes életciklusában meg kell valósítani és biztosítani kell

a) az elektronikus információs rendszerben kezelt adatok és információk bizalmassága, sértetlensége és rendelkezésre állása, valamint

b) az elektronikus információs rendszer és elemeinek sértetlensége és rendelkezésre állása zárt, teljes körű, folytonos és kockázatokkal arányos védelmét.

6. § Az elektronikus információs rendszernek az 5. §-ban meghatározott feltételeknek megfelelő védelme körében a szervezetnek külön jogszabályban előírt logikai, fizikai és adminisztratív védelmi intézkedéseket kell meghatároznia, amelyek támogatják:

a) a megelőzést és a korai figyelmeztetést,

b) az észlelést,

c) a reagálást,

d) a biztonsági események kezelését."

Mindezt a közigazgatási szervezet vezetőjének és az általa kijelölt információbiztonsági felelősnek is feladatul szabja meg, az alábbiak szerint:

„11. § (1) A szervezet vezetője köteles gondoskodni az elektronikus információs rendszerek védelméről a következők szerint: […]

i) gondoskodik az elektronikus információs rendszer eseményeinek nyomon követhetőségéről,

j) biztonsági esemény bekövetkezésekor minden szükséges és rendelkezésére álló erőforrás felhasználásával gondoskodik a biztonsági eseményre történő gyors és hatékony reagálásról, és ezt követően a biztonsági események kezeléséről,

m) felelős az érintetteknek a biztonsági eseményekről és a lehetséges fenyegetésekről történő haladéktalan tájékoztatásáért […]."

„13. § (2) Az elektronikus információs rendszer biztonságáért felelős személy felel a szervezetnél előforduló valamennyi, az elektronikus információs rendszerek védelméhez kapcsolódó feladat ellátásáért. Ennek körében:

a) gondoskodik a szervezet elektronikus információs rendszereinek biztonságával összefüggő tevékenységek jogszabályokkal való összhangjának megteremtéséről és fenntartásáról […]."

A rendelet az 1. táblázatban látható módon részletezi az egyes tevékenységeket is. Azonban ebből az ábrából az is kiderül, hogy az incidenskezelés nem minden közigazgatási szervezetnek törvényben leírt kötelezettsége. Az Ibtv. alapján a törvény hatálya alá tartozó szervezeteket úgynevezett biztonsági osztályokba kell sorolni, amely utal az általuk végzett tevékenységek kritikusságára. A legtöbb szervezet az ötfokozatú skálán 1. vagy 2. biztonsági osztályba sorolandó. Esetükben a biztonsági események kezelése nem kötelező előírás, ez csak a 3. biztonsági osztálytól válik előírttá. Esetükben ad hoc módon történik meg az incidensmenedzsment, azaz amennyiben incidenst észlelnek, azt legjobb tudásuk szerint hárítják el, nem feltétlenül figyelnek az incidens utáni tevékenységekre.

1. táblázat

A rendelet által előírt incidenskezelési tevékenységek

60.	3.1.5.	A biztonsági események kezelése					
61.	3.1.5.l.	Biztonsági eseménykezelési eljárásrend	0	0	X	X	X
62.	3.1.5.2.	Automatikus eseménykezelés	0	0	0	0	X
63.	3.1.5.3.	Információ korreláció	0	0	0	0	X
64.	3.1.5.4.	A biztonsági események figyelése	0	0	X	X	X
65.	3.1.5.5.	Automatikus nyomonkövetés, adatgyűjtés és vizsgálat	0	0	0	0	X
66.	3.1.5.6.	A biztonsági események jelentése	0	0	X	X	X
67.	3.1.5.6.2.	Automatizált jelentés	0	0	0	X	X
68.	3.1.5.7.	Segítségnyújtás a biztonsági események kezeléséhez	0	0	X	X	X
69.	3.1.5.7.2.	Automatizált támogatás	0	0	0	X	X
70.	3.1.5.8.	Biztonsági eseménykezelési terv	0	0	X	X	X
71.	3.1.5.9.	Képzés a biztonsági események kezelésére	0	0	X	X	X
72.	3.1.5.9.2.	Szimuláció	0	0	0	0	X
73.	3.1.5.9.3.	Automatizált képzési környezet	0	0	0	0	X
74.	3.1.5.9.4	A biztonsági események kezelésének tesztelése	0	0	0	X	X
75.	3.1.5.9.4.1.	Egyeztetés	0	0	0	X	X

Forrás: 41/2015. (VII. 15.) BM rendelet

De még a kisebb szervezeteknél sem lehet teljesen negligálni ezt a tevékenységet, különösen akkor nem, ha a szervezet személyes adatokat kezel. Ez pedig a legtöbb közigazgatási szervezetnél előfordul. Esetükben is érvényes ugyanis az Európai Parlament és a Tanács (EU) 2016/679. számú rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről (általános adatvédelmi rendelet), amelyet röviden csak GDPR-ként ismerünk. Ennek 2. szakasza az adatbiztonságról szól, és előírja az incidenskezelési folyamat meglétét, valamint a bekövetkezett, személyes adatokat érintő incidensek esetén az illetékes hatóság, Magyarország esetében a Nemzeti Adatvédelmi és Információszabadság Hatóság értesítését az alábbiak szerint:

„32. cikk: *Az adatkezelés biztonsága*

(1) Az adatkezelő és az adatfeldolgozó a tudomány és technológia állása és a megvalósítás költségei, továbbá az adatkezelés jellege, hatóköre, körülményei és céljai, valamint a természetes személyek jogaira és szabadságaira jelentett, változó valószínűségű

és súlyosságú kockázat figyelembevételével megfelelő technikai és szervezési intézkedéseket hajt végre annak érdekében, hogy a kockázat mértékének megfelelő szintű adatbiztonságot garantálja, ideértve, többek között, adott esetben: […]

b) a személyes adatok kezelésére használt rendszerek és szolgáltatások folyamatos bizalmas jellegének biztosítását, integritását, rendelkezésre állását és ellenálló képességét;

c) fizikai vagy műszaki incidens esetén az arra való képességet, hogy a személyes adatokhoz való hozzáférést és az adatok rendelkezésre állását kellő időben vissza lehet állítani […].”

„33. cikk: *Az adatvédelmi incidens bejelentése a felügyeleti hatóságnak*

(1) Az adatvédelmi incidenst az adatkezelő indokolatlan késedelem nélkül, és ha lehetséges, legkésőbb 72 órával azután, hogy az adatvédelmi incidens a tudomására jutott, bejelenti az 55. cikk alapján illetékes felügyeleti hatóságnak, kivéve, ha az adatvédelmi incidens valószínűsíthetően nem jár kockázattal a természetes személyek jogaira és szabadságaira nézve. Ha a bejelentés nem történik meg 72 órán belül, mellékelni kell hozzá a késedelem igazolására szolgáló indokokat is.

(2) Az adatfeldolgozó az adatvédelmi incidenst, az arról való tudomásszerzését követően indokolatlan késedelem nélkül bejelenti az adatkezelőnek.

(3) Az (1) bekezdésben említett bejelentésben legalább:

a) ismertetni kell az adatvédelmi incidens jellegét, beleértve – ha lehetséges – az érintettek kategóriáit és hozzávetőleges számát, valamint az incidenssel érintett adatok kategóriáit és hozzávetőleges számát;

b) közölni kell az adatvédelmi tisztviselő vagy a további tájékoztatást nyújtó egyéb kapcsolattartó nevét és elérhetőségeit;

c) ismertetni kell az adatvédelmi incidensből eredő, valószínűsíthető következményeket;

d) ismertetni kell az adatkezelő által az adatvédelmi incidens orvoslására tett vagy tervezett intézkedéseket, beleértve adott esetben az adatvédelmi incidensből eredő esetleges hátrányos következmények enyhítését célzó intézkedéseket.

(4) Ha és amennyiben nem lehetséges az információkat egyidejűleg közölni, azok további indokolatlan késedelem nélkül később részletekben is közölhetők.

(5) Az adatkezelő nyilvántartja az adatvédelmi incidenseket, feltüntetve az adatvédelmi incidenshez kapcsolódó tényeket, annak hatásait és az orvoslására tett intézkedéseket. E nyilvántartás lehetővé teszi, hogy a felügyeleti hatóság ellenőrizze az e cikk követelményeinek való megfelelést.”

Az egyes országok kiberbiztonsági felkészültségének mérése összetett feladat, amely számos módszertannal történhet. Ezek közül a felelős hatóságok felé bejelentett kiberbiztonsági incidensek száma és ezek csoportosítása a legszemléletesebb. Magyarország esetében az Ibtv. kötelezővé teszi a biztonsági események jelentését, így ennek a mutatónak a változásából lehetne arra következtetni, hogy egyrészt milyen hatásfokkal készültek fel a szervezetek a kiberbiztonsági kihívások kezelésére, másrészt rendelkeznek-e olyan védelmi képességekkel, amelyek lehetővé teszik a káros események észlelését. A bejelentett események típusainak eloszlása bemutatja, hogy mennyire képesek a magyar közszolgálat szervezetei a komplexebb informatikai támadások felderítésére, egyben következtetni lehet

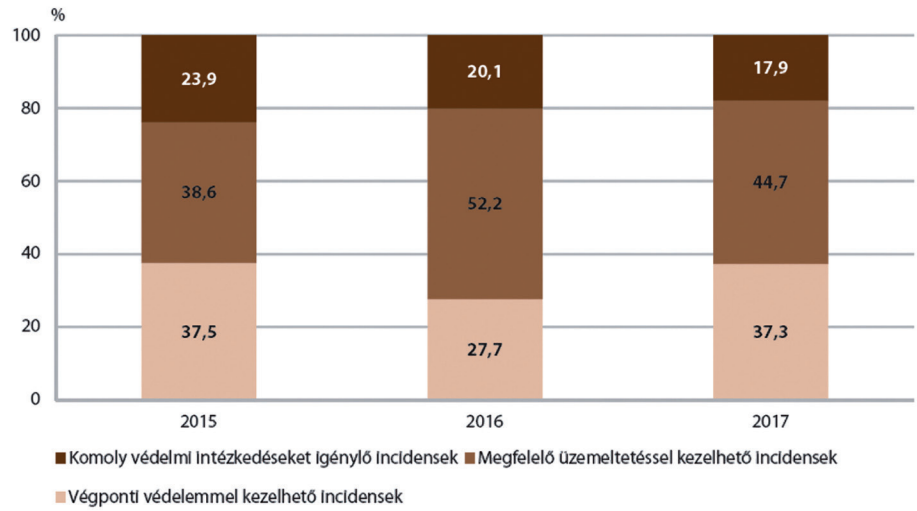
arra, milyen technológiai háttér és mekkora humán erőforrás áll rendelkezésre a magyar kibervédelemben.

A 2017-es statisztikai adatok annyiban újdonságot jelentenek a korábbi adatközlésekhez képest, hogy ezúttal a bejelentett incidensek számát is nyilvánosságra hozta a Nemzeti Kibervédelmi Intézet, amely az Ibtv. alapján elsődlegesen fogadja az adatokat. Eszerint összesen 970 kiberbiztonsági incidens történt a magyar közigazgatásban, azaz ennyi incidensről adott tájékoztatást az a mintegy 5000 szervezet, amely az Ibtv. hatálya alá tartozik. Igaz, hogy közülük nem mindenki tartozik a jelentésköteles kategóriába, mert biztonsági osztályba sorolása 1. vagy 2. A viszonylag alacsony szám azt jelenti, hogy az intézmények többségében valószínűsíthetően az Ibtv. hatálybalépése után 5 évvel is alacsony szinten áll az információbiztonsági felkészültség, nem épült ki az alapvető incidensmenedzsment-képesség sem.

Az incidenstípusok az elhárítás szempontjából három csoportra bonthatók. Az első csoportba a legalapvetőbb védelmi technikákkal kezelhető incidensek tartoznak, ezeket a *végponti védelemmel kezelhető incidensek* kategóriába soroljuk. Ezek 2017-ben az összes jelentett incidens 37%-át tették ki, 9%-os növekedést mutatva az előző évhez képest. A növekedés oka az, hogy a káros szoftverekből eredő biztonsági események jelentősen, 14%-kal nőttek meg, köszönhetően annak, hogy ezek a károkozók minden korábbinál összetettebbek, észlelésük egyre nehezebb.

A második csoportba azok az incidensek tartoznak, amelyek a rendszerüzemeltetéshez, jellemzően az intézmény elektronikus szolgáltatásaihoz, szerverkörnyezetéhez kapcsolódnak. Összefoglalva *megfelelő üzemeltetéssel kezelhető incidensek* kategóriának nevezzük ezt. Az itt megmutatkozó incidensek hátterében jellemzően szakértő- és szakértelemhiány mutatkozik. 2017-ben a jelentett incidensek 45%-a tartozott ebbe a körbe, szemben az előző évi 52%-kal. A közigazgatási informatika erőteljes centralizációja, ezen belül is a NISZ Zrt. kiberbiztonság területén történő képességfejlesztése hosszú távon is hozzájárulhat ahhoz, hogy ez a részmutató csökkenő tendenciát mutasson.

A harmadik csoportba tartoznak azok a támadások, amelyek általában nehezen észlelhetők, kezelésükhöz jelentős beruházás szükséges, ezért *komoly védelmi intézkedéseket igénylő incidensek* néven soroljuk be ezeket. Az iparági statisztikák szerint ezek okozzák a legkomolyabb károkat, sokszor mégis észrevétlenek maradnak. 2017-ben a jelentett incidensek 18%-a tartozott ebbe a körbe, ezzel érdemben nem volt változás a 2016-ban mért 20%-hoz képest. A célzott támadások ezen belül 1%-ot tesznek ki, amely 7 ilyen típusú bejelentett incidenst jelent. Ez továbbra is igen alacsonynak mondható, feltehetően ennél jóval több nem észlelt vagy nem bejelentett célzott támadás történik a magyar közigazgatásban.



3. ábra

Kibervédelmi incidensek aránya a magyar közigazgatásban

Forrás: Kaiser 2018

A cél az, hogy összességében nőjön az incidensek száma, hiszen ez bizonyítja, hogy egyre több szervezet képes észlelni a problémákat, és azokat jelenti is a jogszabályi kötelezettségek szerint. Hosszú távon az ideális az lenne, hogy a végponti védelemmel kezelhető incidensek és a megfelelő üzemeltetéssel kezelhető incidensek kategória aránya csökkenjen a komoly védelmi intézkedéseket igénylő incidensek arányához képest. Ez azt jelentené, hogy Magyarország közigazgatásának kibervédelmi képességei érdemben javultak, hiszen a rövid távon megoldható intézkedéseket megtette, ezek hiánya egyre kevesebb incidenst okoz, miközben kiépültek azok az infrastruktúrák, amelyek a potenciálisan komoly hatású incidensek észlelését teszik lehetővé. Minél többet költ egy kormányzat erre a területre, annál inkább képes lesz észlelni a nemzet biztonságát fenyegető, leggyakrabban rejtve maradó támadásokat.[2]

Információbiztonsági incidensek elemzése egy példa bemutatásával

Egy szervezet információbiztonságának megteremtéséhez többszintű védelmi szolgáltatásokat kell nyújtani. E szolgáltatási portfólió egyik eleme az úgynevezett tűzfalvédelem, amely az internet egészéről érkező forgalom szűrésével képes megakadályozni a káros támadásokat. Az internetes „zaj” kiszűrésével, amely a mesterséges forgalomgeneráló automatáktól származó adatfolyam kizárását jelenti, a szervezetet kiszolgáló számítógépes

[2] Kaiser 2018.

infrastruktúra terhelése is csökken. Ez mind személyes, mind gépi erőforrás-megtakarítást és hatékonyságnövekedést jelent.

Példaesetünkben egy ilyen, nagy hálózati forgalmat bonyolító szervezet tűzfalforgalmát elemezzük. A részletes eredmények ismertetése előtt azonban szükséges valamilyen szinten a háttérben működő technológia fogalmait áttekinteni.

Számítógépek hálózata: az internet

Az információk túlnyomó részét manapság elektronikus formában tároljuk. Ezeket a tárhelyeket és adatfeldolgozó egységeket adattovábbításra alkalmas hálózatok kötik össze. Az egyik ilyen adattovábbítási lehetőség az internet. Itt az egyes számítógépeket vagy szolgáltatási egységeket IP-címek segítségével lehet elérni, a különböző IP-címek közötti adattovábbításért a routerek a felelősek. Ez utóbbi egységek állítják be, hogy egy adott gépről induló adat a célgép IP-címének ismeretében milyen irányban haladjon tovább: vagy közvetlenül a célhoz, vagy egy másik routerhez, amely a megfelelő irányban van.

Minden IP-cím hierarchikus szerkezetben van felépítve. A jelenlegi szabvány, az IPv4-es szabvány minden címet 4 byte-ként ír le, amelyet szokásosan négy egymást követő számmal írunk le, a számok közé pontot helyezve. Ilyen például a 192.168.23.2 IP-cím, ahol minden szám a 0-tól 255-ig terjedő intervallumban fordulhat elő. Ez a címzés azért hierarchikus, mert az utolsó szám a helyi, intézményi vagy egyéb szempontból közelinek számító gépeket különbözteti meg egymástól. Hátulról a második szám már magasabb szintű egységeket különböztet meg. Például egy nagy egyetem esetén a negyedik szám egy tanszék gépeit nevezi meg, a harmadik szám már a karokat vagy intézeteket. A második és az első szám még nagyobb egységekre bontja az interneten elérhető gépek elérhetőségeit. Érdekességként megjegyezzük, hogy a számítógépes eszközök robbanásszerű elterjedésével az IPv4-es címzési szabvány ma már nem ilyen tiszta szerveződésű. Olyan intézmények vagy országok, amelyek már régóta használják az internetet, tipikusan széles címtartományt használnak (például USA, de Magyarország is), míg a frissen feltörekvő államok (például Kína) csak a maradék, szűkös címtartományt használhatják.

A számítógépek az adatokat, illetve a különböző szolgáltatásokat (webhely, e-mail, társalgás, távoli elérés stb.) egy címke segítségével választják szét egymástól. Ez a címke az úgynevezett *port,* amelyre az egy számítógépen belül elkülönítendő forgalmat küldi egy másik gép. Ilyen tipikus port például a 80-as vagy a 443-as, amelyek a titkosítatlan, illetve a biztonságos www-forgalom portszámai. Két számítógép között létrejövő adatkapcsolat során a kapcsolatot kezdeményező gép a saját oldalán kijelöl egy véletlenszerűen kiválasztott portszámot, és innen indítva küldi az adatokat a célgép IP-címére az adatforgalmat fogadni képes szolgáltatásnak megfelelő fogadó portszámmal megcímkézve. Egy ilyen küldő-fogadó IP-port–IP-port párost socketnek nevezünk. A kevés IP-címet használók belső hálózatokat építenek ki tűzfalak mögött, és ezek követik a fenti hierarchikus szerveződést. Természetesen ez a belső IP-tartomány a világ számára kívülről nem látható.

A technikai háttér bemutatását zárjuk azzal a megjegyzéssel, hogy az internetes forgalom úgynevezett csomagalapú adattovábbítási szabványokon alapul. Egy analóg példa segítségével érthetjük meg egyszerűen ennek a lényegét. A hagyományos, nem csomagalapú adattovábbításban (például régi vonalas telefon, CB- vagy FM-rádió) az információ úgy

áramlik, mint amikor beszédben kommunikálunk egymással: folytonos rezgésekbe kódolva megy az adat, és a zavartalan továbbításhoz egy bizonyos sávszélességet le kell foglalni egy közlési csatorna számára. Ezért szólnak a különböző rádióadók különböző frekvenciákon. Ezzel szemben a csomagalapú adatforgalom a borítékba zárt levélhez vagy csomagküldéshez hasonlítható: egy adag információt összegyűjtünk, ezt becsomagoljuk, majd feladjuk a postán a másik címére. Az internetes adattovábbítás során is ez történik: az átküldendő fájlokat feldarabolják kis egységekre, ellátják szabványos fejléccel, majd így küldik egyik gépről a másikra. Különböző adatforgalmak csomagjai összekeveredhetnek, és egymásba fűződve utazhatnak a kábelekben. Mivel egy-egy csomag olyan kicsi, és olyan gyorsan váltogatják egymást az egyes csomagok, a legtöbb internetes szolgáltatásnál nem jelent gondot a csomagosítás.

Ha egy vállalatnak küldünk csomagot, akkor a címzésen feltüntetjük a vállalat nevét, a város és utca/házszám adatokat. A címzésnek ez a része felel meg az IP-címnek. A küldött csomagot valószínűleg valamilyen probléma vagy adott ügyintézés miatt küldjük, ezért a címzésben azt is feltüntetjük, melyik ügyosztálynak szánjuk a küldeményt. Ez felel meg a portszámnak a célgépen. A mi nevünk és postacímünk felel meg a küldő gép IP-címének. Hivatalos küldeménynek szokott lenni valami helyi nyilvántartási száma, amely alapján visszakereshető, vagy az esetleges válasz beazonosítható. Ez az adat felel meg a socket-küldő portszámrészének.

Néhány szó a tűzfalakról

Tűzfalnak nevezzük azt az egységet (amely lehet szoftver egy személyi számítógépen, vagy lehet egy különálló eszköz egy nagyvállalati környezetben), amely elválasztja az intézményi infrastruktúrát a külső gépektől. Az intézményen belüli eszközök általában kevésbé védettek, nagyobb bizalommal kezelik az egymástól érkező információkat. Sok esetben ezek a gépek egyetlen egységes konfiguráció alapján működnek. Ezzel szemben a külső, a világháló oldalán többnyire ismeretlen fenntartású, ismeretlen célból üzemeltetett gépek működnek. Ezek nagy része jóindulatú, valamilyen külső intézmény vagy más ország szolgáltatásait látja el, de vannak kifejezetten rosszindulatú, kártékony céllal működő egységek is. A kártékony gépek felderítése, kiiktatása sokszor komoly felkészültséget, nagy költséget és nemzetközi intézkedéseket igényelne. Az internet „megtisztítása” tehát eleve reménytelen feladat, a saját oldalon kell kiépíteni a megfelelő védettséget.

Ennek a védelemnek az első szintű, a „tűzvonalhoz” legközelebb eső eleme a tűzfal. Ez nem más, mint egy speciális forgalomszűrő szoftver, amely a rajta áthaladó forgalmat folyamatosan monitorozza, és a szabályrendszerben meghatározott, tiltottnak ítélt adatcsomagokat kiemeli. Ezek a csomagok nem jutnak el tehát a célgéphez, a feladójuk úgy érzékeli, hogy a fogadói oldal vagy nem létezik/nem működik (csomagok eldobása esetén), vagy kifejezett visszautasítást kap, azaz „vissza a feladóhoz” jelzéssel kap egy értesítést, hogy az elküldött csomagot biztonsági okokból nem kézbesítették.

Az internetes adatforgalom egy technikai szolgáltatás, amely működtetéséhez különféle eszközöknek szabványos megállapodások szerint kell működnie. Azonban minden technikai eszköz meghibásodhat, vagy átmeneti zavarok léphetnek fel a működésében. Ezért előfordulhat olyan eset, amikor a csomagforgalomban elvesznek egyes küldemények.

Alapesetben ezeket a csomagvesztési problémákat az adatkezelő szabványok megfelelően megoldják, azonban éppen ezeknek a problémakezelő módszereknek a rosszindulatú kihasználásával támadást is lehet intézni egy gép ellen. A tűzfalaknak tehát nemcsak egyszerűen a feladó/fogadó megbízhatóságát kell ellenőrizniük, hanem olyan összetett forgalmi mintázatokat is figyelniük kell, amelyek alapján elkülöníthetők a technikai és a kártékony csomagpótlási esetek.

Végül megemlítjük még azt a tűzfalfunkciót is, amely egy álcát tart a védendő eszközök elé. Ez alatt azt kell érteni, hogy a védendő gépeket olyan mértékben el kell választani a külvilágtól, hogy azok címei külső szereplő számára ne is legyenek láthatók. Ilyenkor a belső gépről érkező forgalom IP-port-számait a tűzfal átírja egy kívülről láthatónak kategorizált IP-port-párosra, de megjegyzi, hogy az erre válaszul érkező adatokat melyik belső IP-címmel rendelkező gép melyik portjára kell küldenie.

Példaadatsor: egy nagy hálózati forgalmú szervezet tűzfalának logfájlja

Egy megfelelően konfigurált és elegendő teljesítménnyel rendelkező tűzfal a rajta átmenő teljes forgalomról képes elmenteni adatokat későbbi elemzés céljára. Egy magas szintű tűzfal esetén ez óriási mennyiségű adatot jelent. Esetünkben 14 órányi működés során közel 760 millió bejegyzés történt. A tűzfal az előző alfejezetben ismertetett feladatainak megfelelően az átmenő forgalmat több kategóriára bontja, amelyeket az alábbi táblázatban ismertetünk.

2. táblázat

A vizsgált nagy hálózati forgalmú szervezet tűzfaláról származó adatforgalmi kategóriák

id	name
1	RT_FLOW_SESSION_CREATE
2	RT_FLOW_SESSION_CLOSE
3	RT_FLOW_SESSION_DENY
4	RT_SCREEN_TCP
5	FLOW_REASSEMBLE_SUCCEED
6	FLOW_REASSEMBLE_FAIL

Forrás: a szerzők szerkesztése

A hat kategóriából az első kettő normális működést ír le: egy adatkapcsolatnak létre kell jönnie, majd le kell zárulnia.

A harmadik kategóriától kezdődnek a problémás esetek. Az RT_FLOW_SESSION_DENY azokat a kapcsolatfelvételi próbálkozásokat jelöli, amelyek korábbi tapasztalatok vagy más forrásokból származó információk alapján potenciális veszélyforrást jelentenek, és tiltva lettek a tűzfalon. Ezekben az esetekben a kapcsolatfelvételként küldött csomagokat a tűzfal eldobja, a feladó semmilyen választ nem kap a kérésére.

A negyedik kategória egy különleges támadási mintázatot jelöl. Az RT_SCREEN_TCP címkével ellátott forgalom úgynevezett SYN FLOOD támadásra utal. Itt egy olyan álságos

kapcsolatfelvételi próbálkozást küld a támadó gép, amely folyamatosan azt jelzi a kiszolgáló szerver felé, hogy nem sikerül felvenni a kapcsolatot, próbáljon újra kapcsolatot építeni. A legáltalánosabban használt, TCP-alapú szabványos protokollok nincsenek felkészítve az ilyen rosszindulatú kérelmek kezelésére, ezért ha a védelmi rendszer nem állítja le vagy nem veszi át e kérelmek kezelését, a kiszolgáló szerverek pillanatok alatt túlterhelődnek, és nem lesznek képesek kiszolgálni a valódi adatigényléseket. Itt érdemes megjegyeznünk, hogy az egyszerű tűzfalvédelmen túl igen összetett, egyéb védelmi rendszerek is léteznek. Ezek tárgyalásától azonban most eltekintünk.

Az ötödik és a hatodik kategóriák a csomagvesztési események kezelésével kapcsolatos bejegyzések. Az egyik azokat az eseményeket rögzíti, amikor a csomagküldési folyamban fellépő zavar valószínűleg valamilyen véletlen vagy nem rosszindulatú okra vezethető vissza, és az átmeneti hiba után a kapcsolatot sikeresen helyre lehetett állítani (FLOW_REASSEMBLE_SUCCEED). A másik esetben azonban a probléma komolyabb, és a kapcsolatot nem sikerült helyreállítani (FLOW_REASSEMBLE_FAIL).

A vizsgált tűzfal a védendő tartományt és a külvilágot kétféleképpen is szétválasztja. Az egyik típusú szétválasztás konfigurációs, szoftveres beállítás, a másik hardveres alapú. A szoftverszintű szétválasztás IP-cím-tartományok szerint (illetve ennél finomabban hangolt szabályrendszer szerint) különbözteti meg, hogy melyik cím tartozik a belső, megbízható (trusted) zónához, és melyik a külső, azaz nem biztonságos (untrusted) zónához. A hardveralapú szétválasztásnál azt használhatjuk ki a forgalom jellemzésére, hogy a tűzfal szerepét betöltő gép két különböző internetcsatlakozójából melyiken érkezett a forgalom.

A tűzfal forgalma alapján definiált hálózatok

A tűzfalon áthaladó forgalom hálózatos vizsgálatában természetes módon irányított, súlyozott hálózatot lehet definiálni. A forgalomban ugyanis mindig van egy küldő oldal és egy fogadó oldal, ráadásul ezek a forrás-cél párosok szinte korlátlan mennyiségben újra előfordulhatnak.

A forgalom topológiáját ábrázoló hálózat esetén az alábbi főbb lehetőségek kínálkoznak a felbontás részletessége szerint csökkenő sorrendben:

- IP/port-IP/port
- IP-IP/port
- IP-IP
- subdomain-subdomain

Minden esetben az éleket – irányukkal és súlyukkal együtt – az határozza meg, hogy melyikről indult és melyikre érkezett a forgalom, és milyen sokszor. Az első, legrészletesebb esetben a hálózatot a forgalmat leíró socket által meghatározott élekből építjük. A második esetben a küldő oldalon nem vesszük figyelembe, hogy milyen portról indult a forgalom, de a fogadó oldalon igen. Ennek azért van külön jelentősége, mert ha egy igénylő-kiszolgáló hálózatot szeretnénk elemezni, akkor a küldő oldalon véletlenszerűen választott portszámoknak nincs jelentősége, míg a fogadó oldalon a különböző szolgáltatásokat vagy funkciókat megkülönböztető portszámokat célszerű külön csúcsokként kezelni.

A harmadik esetben még átfogóbb képet kaphatunk, ha csak a gép-gép, azaz IP-cím és IP-cím közötti kapcsolatokat nézzük. Modern, skálázható megoldások esetén sokszor egyetlen IP-cím forgalmát több, független gép szolgálja ki. De elemzésünkben ez az aspektus nem játszik szerepet.

Az utolsó esetben, amely akár több alesetre is bontható, a magasabb hierarchiaszintű IP-tartományok közti hálózatot vizsgálhatjuk. Megjegyezzük, hogy az IP-tartományok összevonásának a második hálózati típus esetén is lehet értelme, tehát azt kell vizsgálni, hogy egy adott szolgáltatást (célportszámmal leírhatóan) milyen IP-tartományok tudnak nyújtani, vagy milyen IP-tartományok vesznek leginkább igénybe.

Alapvető hálózatjellemzők

Az elemzést kezdjük egy áttekintő jellegű képet adó statisztikai vizsgálattal, amelyből kiderül, milyen méretű hálózatokkal lesz dolgunk. A vizsgált időszak alatt összesen 2 134 769 IP-cím fordult elő a logfájlokban. Ezek egy része külső cím, másik része belső cím. A tűzfal szerepének megértéséhez először szét kell választanunk a belsőnek és a külsőnek számító gépeket, illetve az adatküldőként és adatfogadóként szereplő címeket. Erről nyújt áttekintést a 3. táblázat.

3. táblázat

Az IP-címek és az IP-port-párok száma a küldő/fogadó szerepkörönként leszámlálva

Csúcs vagy él	**darab**
Összes IP	2 134 769
Küldő oldali IP	1 716 399
Fogadó oldali IP	447 235
Küldő IP-port	71 967 874
Fogadó IP-port	106 908 636
Küldő IP – fogadó IP	88 371 569
Küldő IP – fogadó IP-port	174 026 431
Küldő IP-port – fogadó IP-port	392 799 219

Megjegyzés: Egy „fogadó” szerepű IP-cím lehet akár a külső vagy a belső zónában is, és ugyanígy a „küldő” IP is.

Forrás: a szerzők szerkesztése

A 3. táblázat adatai alapján látható, hogy az erős adatforgalom topológiai szempontból is nagy hálózatokat definiál. Mielőtt további statisztikai jellemzőket kezdünk vizsgálni, a tűzfal előző fejezetben ismertetett címkéi és zónái szerint is érdemes áttekintenünk az adatok mennyiségét.

A hat forgalmi címke és a két zóna, illetve a küldő/fogadó szerepkör nagyon sok kombinációs lehetőséget adna, de amint az alábbi táblázatban látjuk, a lehetőségeknek csak kis hányada valósul meg.

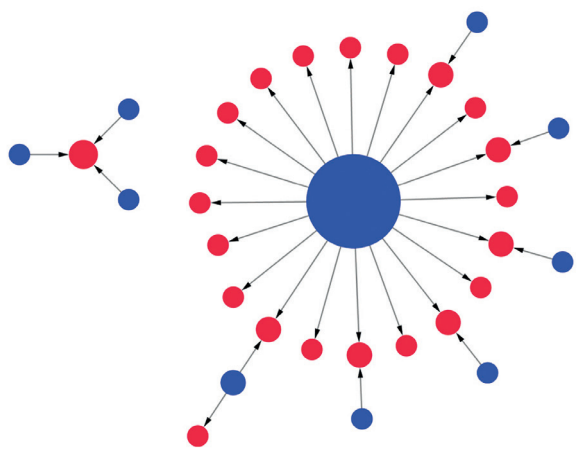
4. táblázat
A forgalom nagysága az 1. táblázatban felsorolt típusok szerint

type	su	st	du	dt	r0	r1	spc	sic	dic	dpc	s3c	s2c	s1c	d3c	d2c	d1c
1	0	1	1	0	1	0	155 961 485	1 362	319 569	685 213	228	66	29	149 325	21 683	215
1	1	0	0	1	0	1	35 663 370	411 526	116 208	298 125	127 621	21 736	203	496	3	3
1	1	1	1	1	1	1	3 396	10	22	23	9	7	6	3	1	1
2	0	1	1	0	1	0	156 030 283	1 422	320 337	755 901	252	71	29	149 456	21 720	215
2	1	0	0	1	0	1	35 249 244	406 515	116 160	290 088	124 668	21 615	203	496	3	3
2	1	1	1	1	1	1	3 373	10	22	23	9	7	6	3	1	1
3	0	1	1	0	1	0	2 522 348	326	598	1 049	69	7	5	403	313	104
3	1	0	0	1	0	1	197 248 136	1 358 827	125 330	105 293 253	58 8470	27 179	212	497	4	4
4	1	0	0	0	0	1	18 721	195	16 053	17 169	177	130	65	495	3	3
5	0	0	0	0	0	0	606	485	108	108	395	299	93	82	61	30
6	0	0	0	0	0	0	273	25	270	270	24	24	21	22	3	3

Forrás: a szerzők szerkesztése

A *type* oszlop a 2. táblázat *id* oszlopa. Az *su, st, du, dt* oszlopok jelölik a küldő/fogadó (s/d: source/destination), külső/belső (u/t: untrusted/trusted) kombinációkat, például ahol az *st* 1, ott a küldő szerepű belső zónában lévő egységekről látunk statisztikát. Az *r0* és *r1* jelöli 1-essel, hogy a két fizikai csatlakozón folyó forgalomból melyeken szerepelt az egység. A többi oszlop mutatja, hogy melyik vizsgált egységből hány darab fordult elő. Az *s* kezdetűek a küldő szerepű egységek, a *d* kezdetűek a fogadók. A nevek második betűje jelzi, hogy milyen egység előfordulásait nézzük: *p* jelöli az IP-port-együttest, az *i* a teljes IP-címet, a *3, 2* és *1* számok pedig az IP-cím hierarchiaszintjét. A *c* betű minden oszlop nevében a darabszámra (count) utal. Például a *d2c* oszlop a fogadóoldali, az IP-cím első két számát tartalmazó tartománycímekből mutatja, hogy hány különböző érték fordul elő.

A táblázat harmadik sorában láthatunk egy speciális tulajdonságú beállítást. Az itt szereplő címeknek az adatsor által lefedett időszak alatt megváltoztatták a megbízhatósági besorolását: voltak egyaránt megbízható és nem megbízható zónában is. Az ilyen IP-címek közötti forgalom hálózatát láthatjuk az alábbi ábrán. Két, egymáshoz nem kapcsolódó komponens figyelhető meg: az egyikben van egy domináns központ, a másik egy kis, hurokban összekötött hálózat.



4. ábra

Megbízhatósági zónát váltó IP-címek hálózata

Megjegyzés: az irányított élek a kialakított kapcsolat irányát mutatják.

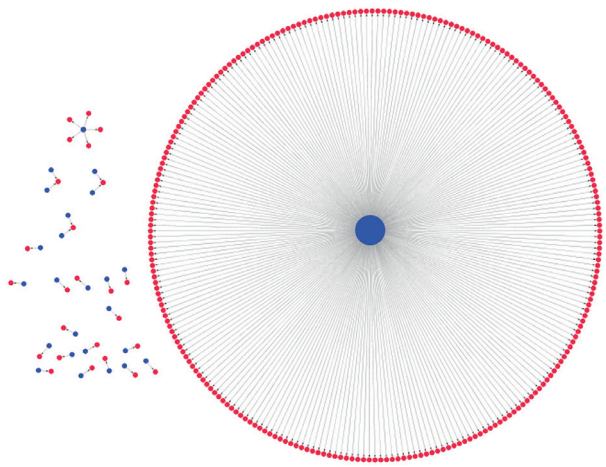
Forrás: a szerzők szerkesztése

Néhány incidensre, illetve automatikusan visszavert támadásra példa

A következőkben néhány incidensre, incidensgyanús esetre, továbbá néhány automatikusan kivédett támadásra látunk példát.

1. *Adatfolyam-megszakadás sikertelen helyreállítással.* Néhány adatkapcsolatot a tűzfal FLOW_REASSEMBLE_FAIL címkével látott el. Ezekben az adatkapcsolatokban valamilyen

probléma miatt megszakadt a kapcsolat. Ilyen eset akkor fordul általában elő, amikor a kapcsolatot létrehozó vagy fenntartó csomagok feltöredeznek, és a csomagtöredékekből nem sikerül helyreállítani az eredeti csomagot. Ha hálózat formájában ábrázoljuk ezeket a megszakadt kapcsolatokat, egy sajátos mintázatot láthatunk.



5. ábra

Megszakadó kapcsolatban részt vevő IP-címpárok hálózata

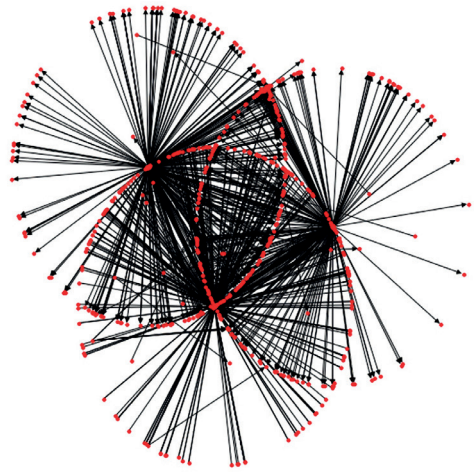
Forrás: a szerzők szerkesztése

A hálózat egyáltalán nem nevezhető véletlenszerűen kiválasztott IP-címek hálózatának. Néhány véletlenszerű pár természetesen van, de ugyanakkor az is feltűnő, hogy egyetlen domináns IP-címnek van igazán sok kapcsolata. Ezekben az IP-párokban a domináns cím egy belső cím, amelyből számos kapcsolat indul ki.

2. *Biztonságos zónából indított, de mégis letiltott adatkapcsolat (FLOW_DENY).* Ennél az incidenstípusnál a kezdeményező fél elvileg egy biztonságos gép, ugyanis a tűzfal belső oldalán tartózkodik. A tűzfal mégis kockázatosnak klasszifikálja a forgalmat, ezért letiltja, és a csomagokat eldobja. Ilyenkor, ha a gép előtt valaki ül, és véletlenül kezdeményezte ezt a típusú kapcsolatot, akkor azt látja, hogy az előtte lévő számítógép rendben működik, de mégis az internettel problémák adódnak, a kis homokóra hiába forog a képernyő közepén, az adatkapcsolat nem épül fel.

A letiltott kapcsolódások forgalmának hálózatos ábrája alapján, amelyet alább láthatunk, feltételezhető, hogy ezeket a tiltott forgalmakat (két kivételtől eltekintve, amelyek önálló, pici komponenst alkotnak) valamilyen automata program okozhatta. Itt tehát egy támadás miatt megsérült/feltört belső gépekre látunk példát, amelyek az incidenskezelési folyamat időigénye miatt átmenetileg még működtek egy ideig. A tűzfalon már nem jutott ki róluk információ, de a belső hálózaton még fennállt a továbbfertőzés veszélye, például külső lemez vagy egyéb adathordozó segítségével. Tehát a mindennapi gyakorlatban, ha azt látjuk, hogy a saját vagy a munkatársunk/beosztottunk gépe le van tiltva az internetről,

szakember nélkül ne kezdjük el gyorsan átvinni az adatokat egy másik gépre. Azzal ugyanis a már felismert fertőzést vihetjük át más gépekre.

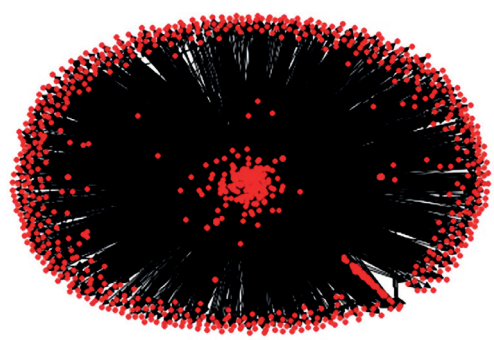


6. ábra

A tűzfal biztonságos, belső oldaláról indított, de mégis letiltott kapcsolatok forgalmának kapcsolati hálózata

Forrás: a szerzők szerkesztése

A 6. ábrán három, egymással átfedést képező csillagalakzatot látunk, amelyek metszetében számos, egymással többszörösen kapcsolódó IP-cím helyezkedik el (lásd: 7. ábra). A sugaras szerkezet azt mutatja, hogy egy-egy gép nagyon sok egyedi külső címhez próbál kapcsolódni. A többszörös kapcsolódások pedig azt sejtetik, hogy lehet valami közös, eredetileg a tűzfalon kívülről származó, közös forrása vagy célszolgáltatása ezeknek az elhalt kapcsolatfelvételi próbálkozásoknak.



7. ábra

A 6. ábra topologikus elrendezésben

Megjegyzés: ez az ábra a többszörösen kapcsolódó csúcsokat szeparáltan ábrázolja a kevés kapcsolódású pontoktól.

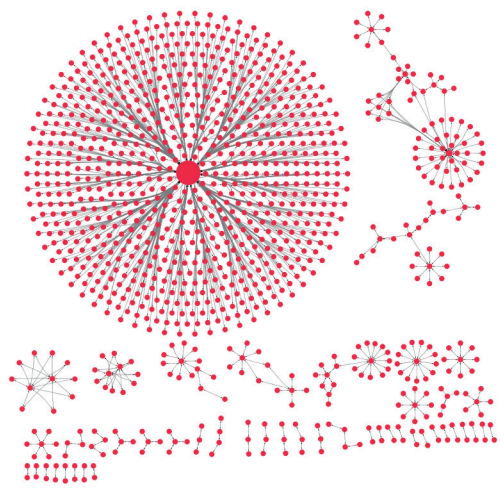
Forrás: a szerzők szerkesztése

3. *Tükörforgalom*. Eddig a gép-gép, azaz IP-IP irányított hálózatokkal szemléltethető potenciális támadást vagy támadást követő incidenshelyzetet láttunk. Ebben a példában egy részletesebb forgalmi hálózatra hozunk példát. A legrészletesebb ábrázolás esetén az IP-címek mellett az igénybe vett szolgáltatás (célgépen a portszám) és a kezdeményező gépen is a portszám külön csúcsként van ábrázolva. Azaz egy él húzódik az IP-port-forrás és IP-port-cél socket mentén.

Fontos megjegyezni, hogy a portszámokat a fogadó és igénylő oldalon általában aszimmetrikusan szokták kiosztani. A szokásos szolgáltatások (például web, e-mail, bejelentkezés) tipikusan alacsony portszámokon történnek (például 80, 443, 25, 993, 22), míg a forgalomkezdeményezés valamilyen magasabb portszámot kap (például 21485, 37245). Ebből kiindulva a legnagyobb felbontású IP-port → IP-port irányított hálózatban nem várunk erős komponenst vagy szimmetrikus éleket. Sőt, inkább egy páros gráfot várunk, ahol két csoportra bonthatók a csúcsok, és az élek sosem kapcsolnak össze csoporton belüli csúcsokat, csupán a csoportok között futnak. Természetesen a gyakorlatban nincs kötelező érvényű feltétel a kezdeményező és a fogadó portszámokra vonatkozóan.

Az előzetes várakozás ellenére a forgalmi adatok feldolgozásakor ebben a részletes hálózatban fel lehetett fedezni egy erősen összefüggő komponenst. Továbbá ebben a hálózatban szimmetrikus éleket is fel lehetett fedezni, azaz voltak olyan socketek, amelyek megfordítva is előfordultak. Azaz vannak olyan esetek, amikor az A gép az x portjáról kiindulva csatlakozott a B gép y portjára, és valamennyi idő elteltével a B gép fog csatlakozni az y portról kiindulva az A gép x portjára.

Az alábbi ábra azon IP-címek hálózatát mutatja, amelyek a fent leírt tükörforgalommal kapcsolatba kerültek egymással.



8. ábra

Tükörforgalommal kapcsolatba került IP-címek kapcsolati hálózata

Forrás: a szerzők szerkesztése

A hálózat szerkezete érdekesen szélsőséges eloszlást mutat. Egy központi IP-cím dominálja a kapcsolatok jelentős részét. Ennek az IP-nek extrém magas a fokszáma, szinte minden

más IP-hez kapcsolódik ebben a hálózatban. Azok az IP-k, amelyekhez nem tudott kapcsolódni, azokkal még áttételes kapcsolatban sincs ez a domináns cím. A hálózat egy óriáskomponensre (a domináns IP holdudvarára) és több, kisebb, szeparált komponensre bomlik. A domináns IP komponense 794 csúcsból áll, míg a méretben következő már csak 68 IP-t fog össze. Ez a komponens viszont már komplexebb szerkezetű, kaszkádszerű terjedési mintázatot mutat. Ezt a kaszkádkomponenst látjuk a domináns csillag mögött a 8. ábrán.

4. Érdemes még megemlíteni két további gyakori incidenstípust: a *port scan* és a *syn flood* támadásokat. Mindkettőről a kapcsolódó BOTA-projekt kapcsán mutatunk példát. A syn flood lényegét a technikai bevezetőben már ismertettük. A port scan támadásról itt csak annyit említünk, hogy ez a támadás ahhoz hasonlítható, mint ahogy a rajzfilmekben szokták a régi idők rablóit ábrázolni. A rabló hord magával egy nagy karikán egy csomó kulcsot, és amikor odaér a bejárati ajtóhoz, nekiáll végigpróbálni az összes nyitási lehetőséget. Ha szerencséje van, akkor talál olyan kulcsot, amely beleillik a zárba, és akkor be tud törni. A port scan esetén a támadó végigpróbálja a lehetséges portszámokat a célgépen, hátha talál egy olyan szolgáltatást, amelyhez ismeri a belépési kódot vagy a nyitható kiskaput.

Felhasznált irodalom

Aash, P. (2017): Incident Response. Validation, Containment & Forensics. *CISO Platform,* 2017. 05. 22. Elérhető: www.slideshare.net/cisoplatform7/incident-response-validation-containment-forensics (A letöltés dátuma: 2019. 03. 12.)

Cichonski, P. – Millar, T. – Grance, T. – Scarfone, K. (2012): *Computer Security Incident Handling Guide.* NIST Special Publication 800-61 Revision 2. Gaithersburg, Maryland, National Institute of Standards and Technology. DOI: http://dx.doi.org/10.6028/NIST.SP.800-61r2

Kaiser T. szerk. (2018): *Jó Állam Jelentés 2018.* Budapest–Pécs, Dialóg Campus Kiadó.

Hivatkozott dokumentumok

2013. évi L. törvény az állami és önkormányzati szervek elektronikus információbiztonságáról.

41/2015. (VII. 15.) BM rendelet az állami és önkormányzati szervek elektronikus információbiztonságáról szóló 2013. évi L. törvényben meghatározott technológiai biztonsági, valamint a biztonságos információs eszközökre, termékekre, továbbá a biztonsági osztályba és biztonsági szintbe sorolásra vonatkozó követelményekről.

Európai Parlament és a Tanács (EU) 2016/679 rendelete (2016. április 27.) a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról, valamint a 95/46/EK rendelet hatályon kívül helyezéséről (általános adatvédelmi rendelet).

Orbók Ákos – Dobos László – Palla Gergely – Pollner Péter

Az egyetemi polgárok wififelhasználói szokásai a Ludovika Campuson

Bevezetés

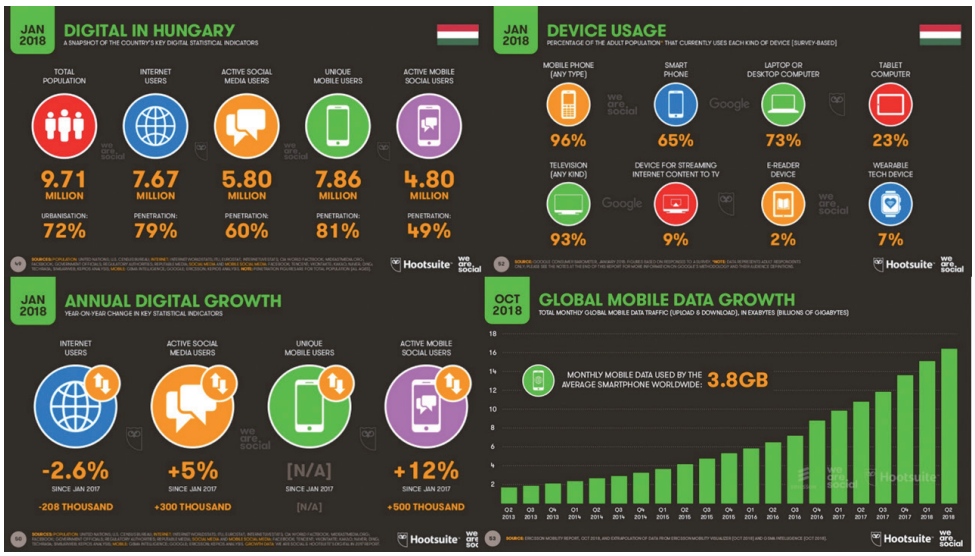
Ebben a tanulmányban röviden összefoglaljuk azokat az adatokat, amelyek Magyarország internethasználati szokásairól ismertek. Majd bemutatjuk a wifihálózatok használatával kapcsolatos adatvédelmi és információbiztonsági elveket és veszélyeket. A kutatás fő célja, hogy feltérképezzük a Nemzeti Közszolgálati Egyetem Ludovika Campusán a wifihálózatot használók szokásait – többek között a hálózattudomány eszközeivel.

Magyarország felhasználói szokásai és eszközei

Magyarországon az internet elterjedtsége a Wearesocial.com kérdőíves felmérései alapján 2018-ban közel 80%-os volt, azaz valamivel az Európai Unió átlaga alatt maradt. Az Eurostat azonos felmérése, amely a háztartások kapcsolódását mérte, 82%-ot mutatott 2018-ban. Az 1. ábrán az internetelterjedtség mellett találhatunk olyan felméréseket is, amelyek a felhasználók eszközeit és szokásait mutatják. A magyarországi adatok mellett egy globális eszköz- és adatforgalom-növekedési grafikont is láthatunk: világosan kiderül, hogy itt általános tendencia a növekedés. Ez a mobil-adatforgalom tekintetében exponenciálisnak mondható.

A magyarországi adatokból az is látszik, hogy a legtöbbet a mobil- és okoseszközöket (azon belül az okostévéket) használják. A hagyományosabbnak mondható számítógépes forgalom másodlagosnak tekinthető. Viszont ha az adatforgalmat tekintjük, ott a hagyományosabb eloszlás lesz jellemző, ahogy az a 2. ábrán is látható. Azonban ott is látszik a számítógép-használat csökkenő tendenciája (a tavalyi felméréshez képest 16%-os a csökkenés) és a mobileszközök alkalmazásának növekedése (az előző évi felméréshez képest 61%-os a növekedés). Tehát a globális adatforgalom-növekedés főleg a mobileszközökön jelentkezik.

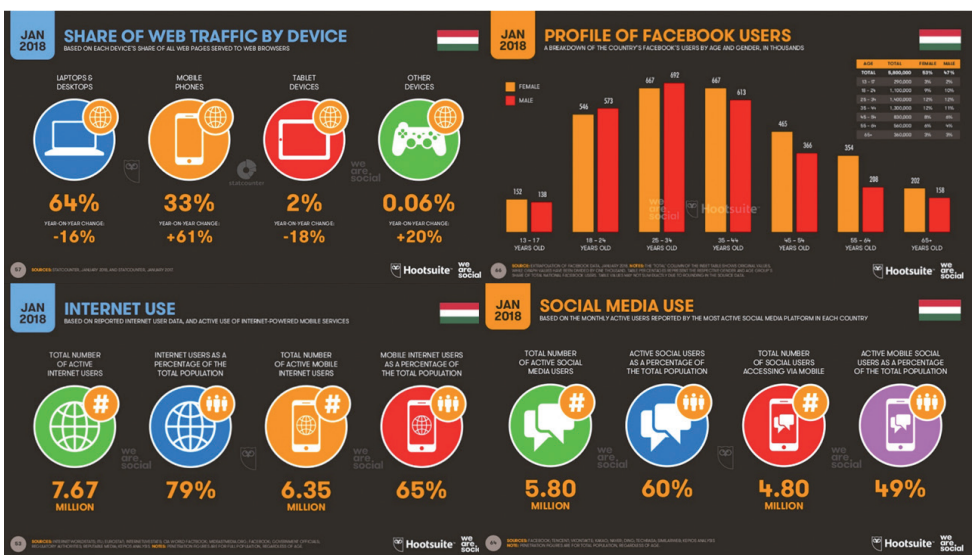
A másik érdekesnek mondható adat a közösségi média használata Magyarországon (2. ábra). Ez ugyanis a legmagasabb egy főre jutó közösségimédia-használat az EU-n belül. A teljes népesség 60%-a használja a közösségi média valamilyen formáját, ezt majdnem teljes mértékben mobileszközökön teszik. A legnagyobb közösségimédia-felületre regisztrálók korcsoportjainak aránya nem mutat nagy eltérést más EU-országétól, a 18–44 éves korosztály van jelen a legnagyobb mértékben.



1. ábra

Magyarország internetfelhasználói szokásai és eszközei

Forrás: Wearesocial.com



2. ábra

Magyarország internetfelhasználói szokásai és a közösségimédia-használat

Forrás: Wearesocial.com

Adatvédelem a wifis adatgyűjtés során

A fejezetben azt vizsgáltuk, milyen szokások jellemzik a Ludovika Campus wifihálózatának felhasználóit. A wifihálózat naplóadatait, amelyeket használhattunk a vizsgálathoz, anonimizálva kaptuk meg, úgyhogy az adatvédelmi incidens esélyét is elkerültük. Azonban felmerül a kérdés, hogy ezek az adatok tekinthetők-e személyes adatoknak, vagy sem. A kérdés megválaszolásához szükségünk van néhány fogalom tisztázására is.

Személyes adat: azonosított vagy azonosítható természetes személyre („érintett") vonatkozó bármely információ. Azonosítható az a természetes személy, aki közvetlen vagy közvetett módon, különösen valamely azonosító, például név, szám, helymeghatározó adat, online azonosító vagy a természetes személy testi, fiziológiai, genetikai, szellemi, gazdasági, kulturális vagy szociális azonosságára vonatkozó egy vagy több tényező alapján azonosítható.

Profilalkotás: személyes adatok automatizált kezelésének bármely olyan formája, amelynek során a személyes adatokat valamely természetes személyhez fűződő bizonyos személyes jellemzők értékelésére, különösen a munkahelyi teljesítményhez, gazdasági helyzethez, egészségi állapothoz, személyes preferenciákhoz, érdeklődéshez, megbízhatósághoz, viselkedéshez, tartózkodási helyhez vagy mozgáshoz kapcsolódó jellemzők elemzésére vagy előrejelzésére használják.

Álnevesítés: a személyes adatok oly módon történő kezelése, amelynek következtében további információk felhasználása nélkül többé már nem állapítható meg, hogy a személyes adat mely konkrét természetes személyre vonatkozik, feltéve, hogy az ilyen további információt külön tárolják, és technikai és szervezési intézkedések megtételével biztosított, hogy azonosított vagy azonosítható természetes személyekhez ezt a személyes adatot nem lehet kapcsolni.

Mikor válik személyes adattá egy technikai adat? – tehetjük fel a kérdést. Az egyszerű válasz az, hogy amikor az adatok alapján meg lehet mondani, kihez tartoznak, tehát azonosítani lehet a személyt. A jogi válasz elég egyértelmű, de műszaki szemszögből nem ennyire egyértelmű a kérdés. Az azonosíthatóság főleg az eszközökre vonatkozik – az adatok pedig az általunk is vizsgált naplóban találhatók. Tehát elsőre azt mondhatnánk, hogy nem kell semmit titkosítani, hiszen az eszközöket bárki használhatja, így nem azonosíthatók vele a felhasználók, tehát nem számít személyes adatnak. Azonban a naplózás során rögzítik az eszközök úgynevezett MAC-címeit is. E cím az eszközhöz tartozó fizikai cím, tehát nem vagy csak nehezen átírható. A MAC-címek segítségével beazonosíthatók az eszközök, és ha valamilyen rosszindulatú tevékenység folytán kikerülnek az eszközökön tárolt adatok, máris személyessé válik az adatbázis.

A következő fejezetben egy olyan rosszindulatú tevékenységet ismertetünk, amely a wifihálózat gyengeségeit használja ki az adatlopásra.

A wifi Pineapple működése és kockázatai

A wifi Pineapple támadás lehetővé teszi a használójának, hogy ellopja az adatokat a nyilvános wifihálózatok használóitól. A médiában a hackereket gyakran maszkrajongó és számítógépes szakemberek keverékeként ábrázolják, akik a billentyűzettel bármilyen digitális

eszközt képesek irányítani. Ezzel a módszerrel szinte bármilyen eszközt, amely az internethez csatlakoztatott, képesek az irányításuk alá vonni, még akkor is, ha nem ismerik az SSL-t egy SSID-ből. Mivel ez annyira olcsó és könnyen használható, fontos megérteni, hogyan működik a Pineapple, hogy képesek legyünk megvédeni magunkat a támadótól.

Ahhoz, hogy érthető legyen a wifihálózat működése és a támadás módja, először meg kell értenünk néhány fogalmat. Az előbb említett SSL és SSID azok a létfontosságú biztonsági eszközök, amelyek szavatolják a wifi kommunikációs csatornán a biztonságos működést.

Az SSL- (Secure Sockets Layer, azaz biztonságicsatlakozó-réteg) tanúsítványok arra szolgálnak, hogy létrejöhessen egy biztonságos, titkosított csatorna a kliens és a szerver között. Bizonyos információk – mint a hitelkártyaadatok, fiókbelépéshez szükséges adatok és egyéb kényes információk – átvitelének titkosítással kell történnie, hogy kizárjuk az adatok kiszivárgását. Az SSL-tanúsítvánnyal adataink titkosításon esnek át, mielőtt azok internetre kerülnének. A titkosított adatot csak a célszerver képes lefordítani. Ez biztosítja, hogy a weboldalon megadott adataink ne kerülhessenek illetéktelenekhez.

Minden vezeték nélküli helyi hálózatnak (WLAN) van egy egyedi hálózatnév-azonosítója. Ez más néven az SSID (Service Set Identifier, szolgáltatáskészlet-azonosító). A wifihálózati adapter beállításakor meg kell adni az SSID-t. Ha létező WLAN-ra szeretnénk csatlakozni, akkor annak a nevét kell használni. Ha saját WLAN-t hoznánk létre, akkor választhatunk nevet, és minden számítógépen ezt a nevet kell használni. A név legfeljebb 32 karakterből állhat, és csak betűket és számokat tartalmazhat. Az SSID vagy hálózatnév a hozzáférési pontban vagy vezeték nélküli routeren adható meg. [1]

3. ábra
Wifi Pineapple
Forrás: https://shop.hak5.org/products/wifi-pineapple

A Pineapple-t 2008-ban a Hak5 nevű cég alkotta meg, amely betöréstesztelő eszközöket, avagy „pentesters” eszközöket fejlesztett. Az ilyen eszközökkel dolgozó vállalkozásokat rendszerint olyan szervezetek kérik fel, hogy támadják meg saját hálózataikat, amelyeknek fontos, hogy felfedezzék a sebezhetőségeket, mielőtt néhány rosszindulatú szereplő

[1] COLE–BARBER 2007.

felfedezné azokat. A Pineapple lehetővé teszi a használójának, hogy könnyen végrehajthasson olyan kifinomult támadásokat a nyilvános wifihálózatokban, amelyekkel megtudhatják, hogyan működnének a támadások, és hogyan védjék meg azoktól a hálózatot.

A Pineapple-ök nem különböznek a szokásos wifihozzáférési pontoktól, amelyek otthoni vagy irodai internetes hozzáférést biztosítanak, csak erősebbek. Több jeladót használnak egyszerre, nem csak egyet, amelyet a legtöbb forgalomirányító megtalál. Ez azt jelenti, hogy egy Pineapple képes egyszerre több eszközzel egyidejűleg működni, nem csak néhány tucattal. Továbbá a Pineapple webes felületét bonyolult hálózati támadások végrehajtására optimalizálták. A Pineapple felbecsülhetetlen értékű eszköz a pentesterek számára, de népszerűsége annak is köszönhető, hogy más műveletekre is felhasználható. A hackerek könnyen kezelhetik az eszközt, így érzékeny személyes adatokat gyűjthetnek a gyanútlan felhasználóktól a nyilvános wifihálózatokon. Ilyen támadásra példa a spoofing, amikor egy hacker egy szolgáltatást vagy eszközt szimulál, hogy hozzáférjen az áldozat adataihoz.

Man-in-the-middle- (MITM-) támadás és az Evil portál

Az MITM-támadás a felhasználó lehallgatásának egy módja azáltal, hogy beillesztik a Pineapple-t a felhasználó eszköze és a legitim wifihozzáférési pontok közé (az adatok hálózaton való átirányításával, nem feltétlenül fizikailag közéjük). A Pineapple ilyen esetben azt állítja magáról, hogy ő a legitim wifihozzáférési pont, így az összes adatot megkapja, mivel továbbítja az adatokat az eszközről a hozzáférési pontra. Ez olyan, mintha egy levelet, amit az ön postaládájába szántak, egy idegen nyitotta volna meg, elolvasta volna, majd visszaküldte volna a postaládájába. Az egész műveletből valószínűleg semmit nem vesz észre az áldozat, hiszen nincs adatvesztés, csak másolás. Gondolhatnánk, hogy nagyon egyszerű védekezni az ilyen támadásokkal szemben: nem kell nyilvános wifire csatlakozni. Ez a stratégia az esetek többségében be is válhat, kivéve, ha a Pineapple egy már általunk rögzített wificsatlakozási pontnak álcázza magát, amelyre az eszközünk automatikusan csatlakozik. Például a wifi és Bluetooth kikapcsolása az iOS 11-ben valójában nem kapcsolja ki a wifi- vagy a Bluetooth-funkciót. Tehát ha valakinek vannak rögzített hálózatok az eszközén, egy ilyen támadással szemben védtelen. Ezután merülhet fel a következő kérdés, hogy hogyan képes a támadó feltérképezni azokat a mentett hozzáférési pontokat, amelyeket a készülék tárol. A Pineapple trükkje az, hogy szkenneli azokat a szolgáltatáskészlet-azonosítókat (SSID) – azaz a wifihálózatok nevét –, amelyek a közelben lévő eszközökről sugároznak. Ugyanis az eszközeink bekapcsolt wifivel állandóan keresik a mentett hálózatokat, ezzel felfedve a behatolási pontokat a Pineapple-nek. Bármikor, amikor telefonjával vagy számítógépével csatlakozik a wifihálózathoz, a készülék menti a wifihálózat SSID-jét, hogy a jövőben gyorsabban tudjon csatlakozni a wifihálózathoz. De ennek a kényelemnek nagy ára van, ahogy láthattuk.

Nézzünk egy gyakorlati példát! Tegyük fel, hogy rákapcsolódik a wifihálózatra az egyetemi könyvtárban, és hálózatát NKE-GUEST-nek nevezik. Miután elhagyta a könyvtárat, telefonja vagy laptopja elkezd sugározni egy olyan jelet, amely alapvetően megkérdezi, hogy a wifihozzáférési pont az eszköz körül az NKE-GUEST-e. Az eszköz minden olyan hálózathoz csatlakozik, amelyhez a múltban csatlakozott. A Pineapple kihasználhatja ezt a funkciót, ha beolvassa a közelben lévő eszközök által sugárzott összes SSID-t. Ezután

ezeket az SSID-ket lemásolja, hogy megpróbálhassa az eszközökkel elhitetni azt, hogy ő egy hozzáférési pont, amelyhez a múltban kapcsolódott. Tehát a fenti példát használva, a Pineapple látni fogja, hogy a telefon megkérdezi: „Ez a hálózat az NKE-GUEST?", majd elkezdi sugározni saját jeleit, amely azt mondja: „Igen, én vagyok az NKE-GUEST, csatlakozz hozzám."

Ez olyan, mintha sétálna egy kulccsal a kezében hazafelé, és megkérdezne minden idegent, akivel találkozik, hogy „Te vagy a szobatársam?". A legtöbb esetben ezek az idegenek azt mondják „nem", de ezzel azt is kockáztatja, hogy egy rosszindulatú idegenhez fordul, aki hazudni fog önnek, és azt mondja: „Igen, persze én vagyok a szobatársad. Kérlek, engedj be", majd ezután ellopja minden értékét.

De az eszközök elérése a Pineapple segítségével csak az MITM-kihasználás egyik fele. A támadónak képesnek kell lennie arra is, hogy elolvassa az adatokat az eszközről a Pineapple-ön keresztül. Ennek több módja van. A Pineapple-t fel lehet használni egy Evil Portal létrehozására, amely alapvetően a weboldalak hamis változatait hozza létre felhasználói nevek, jelszavak, hitelkártya-információk vagy egyéb érzékeny adatok rögzítésére. Erre egy helyi szervert hoznak létre a támadó számítógépén, hogy olyan weboldalt tartson fenn, amely ismerős bejelentkezési oldalnak tűnik, például mint a Gmail vagy a Facebook. Ezek az oldalak egyszerűen duplikálhatók ingyenes online szolgáltatások használatával. Ezután a támadó beállítja a Pineapple-t, hogy amikor minden olyan eszköz, amely hozzá csatlakozik, megpróbál egy olyan webhelyet felkeresni, mint a Twitter vagy a Facebook, akkor valójában átkerül a hamis weboldalra, amelyet a támadó számítógép szolgáltat. Ha az áldozat beírja az adatait ezen az oldalon, mondjuk a felhasználónevét és a jelszavát, akkor azok a támadóhoz kerülnek anélkül, hogy a felhasználó valaha tudta volna, hogy ellopták őket.

Az MITM-támadással kapcsolatos információk begyűjtésének másik módja a Pineapple számára készített modulok használata, amelyek blokkolják a kényszerített HTTPS-titkosítást, és elolvassák azokat az adatokat, amelyek egyébként biztonságban lennének.

A felhasználók HTTPS-verzióra való kényszerítése nagyszerű módja annak, hogy megerősítsék a webhelyek biztonságát, de a böngészés a felhasználó HTTP-kérelmével kezdődik, amelyet kihasználhat a Pineapple. Az SSLSplit nevű modul képes felügyelni a HTTP-kéréseket a felhasználó eszközéről, ha csatlakozik az Pineapple-höz. Ezután ezt a kérést továbbítja a megfelelő kiszolgálóhoz, de amikor a kiszolgáló a biztonságos HTTPS-kapcsolattal válaszol, a Pineapple elrejtheti a biztonságos réteget, és a webhely HTTP-verzióját mutatja a felhasználónak. Ezen a ponton a felhasználó csak a webhely hitelesítés nélküli verziójával találkozik, amely majdnem teljesen megegyezik a hitelesítettel. Az egyetlen különbség az lesz, hogy a kis zárikon eltűnik a címsor bal sarkából.

A támadási példák egyértelműen bemutatják a titkosított kommunikációs protokollok (például HTTPS) fontosságát. Ezek nélkül az összes olyan adat, amely az eszköz és a hozzáférési pont között mozog, könnyen olvasható bárkinek, aki Pineapple-lel rendelkezik.

Hogyan védekezhet az adatlopás ellen?

A fent tárgyalt hackek csak a jéghegy csúcsát képezik. Szerencsére csak néhány egyszerű lépésre van szükség ahhoz, hogy megvédhesse az ember magát attól, hogy adatait ilyen módon ellopják.

A nyilvános wifihálózat kerülése. A legegyszerűbb, ha csak ismerős és megbízható wifihálózatokhoz kapcsolódik eszközével. Az otthoni hálózata például szinte biztosan védve van a Pineapple-támadástól. Ennek az az oka, hogy egy Pineapple-nek is hozzá kell férnie ahhoz a hálózathoz, amelyen a forgalmat figyelni próbálja, ezért ha a támadó nem fér hozzá otthoni wifihitelesítő-adataihoz, akkor nem lesz képes a fent leírt Pineapple-támadásra. Ugyanez érvényes az irodájában is. A Pineapple-támadás valódi veszélye a nyilvános hálózatokon van – olyan helyeken, mint a helyi kávézó vagy a repülőtér; ezek lehetnek a támadás elsődleges helyszínei. A legtöbb ember nem tölti az idejét azzal, hogy ellenőrizze, hogy a „free_airport_wifi" hozzáférési pont legitim-e, általában csak csatlakozik gondolkodás nélkül.

Amikor egy hálózati kapcsolat létrehozásáról van szó, a tudatosság kulcsfontosságú. A legbiztonságosabb lehetőség az, hogy soha ne használjon közvetlenül nyilvános wifihálózatokat. Ez azonban komoly hátrányokkal járhat, főleg kényelmi szempontból.[2]

Virtuális privát hálózatok (VPN). Ha nyilvános wifihozzáférésre van szüksége, akkor az a legjobb, ha VPN-t használ. A VPN egy biztonságos módja a neten való szörfözésnek. A VPN-kiszolgáló titkosítja az adatait, mielőtt az úticél felé irányítja. Lényegében védőhéjat hoz létre az ön adatai számára, ezáltal azok érthetetlenek lesznek a rosszindulatú szemek számára. A támadó láthatja, hogy a készülék csatlakozott a Pineapple-höz, de ha VPN-t használ, akkor nem láthatja az általunk forgalmazott adatokat, csak egy titkosított adathalmazt. Ez vonatkozik a lehallgatókra is – legyen az wifi Pineapple, az internetszolgáltatója, a munkáltatója vagy akár az állam.

Végül fontos, hogy amikor készen áll egy nyilvános wifihálózathoz való csatlakozásra, akkor úgy állítsa be a telefonját vagy számítógépét, hogy az *„felejtse el" a hálózatot.* Így a készülék nem fogja folyamatosan sugározni az eddig csatlakozott hálózatok SSID-jét, amelyet egy Pineapple-támadó meghamisíthat. Sajnos ennek az Androidon vagy az iPhone-on nincs egyszerű módja, és minden hálózatot kézzel kell törölni a telefon beállításaiból (a „hálózatok kezelése" lapon). Egy másik egyszerű megoldás az, ha kikapcsolja a wififunkciót, amikor nem használja – bár ezt már nem olyan könnyű elvégezni néhány eszközön (például az Iphone-on) –, de főleg ne hagyja, hogy a készülék automatikusan csatlakozzon a wifihálózatokhoz. A Pineapple-támadást egyszerűen elkerülheti, csak tudatosan kell használnia a készülékeit. Figyeljen a hálózati beállításokra és az internetes tevékenységére. Az ICT-eszközök manipulálásának minden lehetősége ellenére a hackerek sikeressége továbbra is nagymértékben függ az emberi gondatlanságtól.

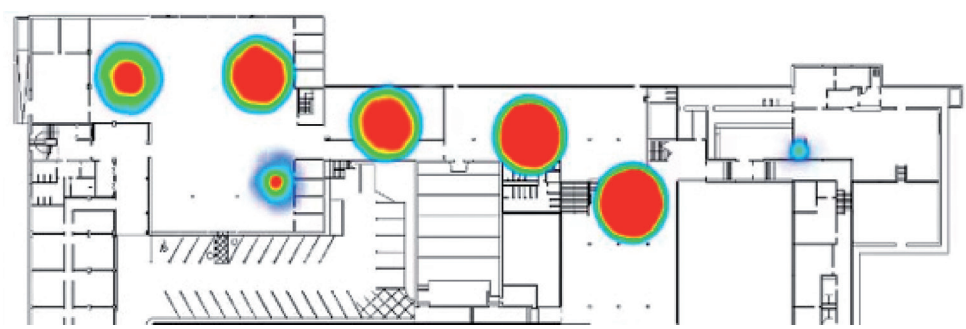
A Madridi Egyetem kampuszán végzett wifihálózat-elemzés eredményei (Smart CEI Moncloa)

A támadások bemutatása mellett fontos, hogy megismerjük, mennyire hasznos lehet, ha a wifihálózatokat a várostervezés vagy a szociális kutatások területén használják. A következőkben egy olyan felmérést mutatunk be, amelyet szintén egy egyetemi kampuszon végeztek el a kutatók a polgárok mozgását vizsgálva.

[2] Oberhaus 2017.

A Madridi Egyetem (Universidad Politécnica de Madrid, UPM) okosváros-programjának (Smart CEI Moncloa) alapja a szenzorhálózaton kívül a wifihálózatra csatlakozók felmérése volt különböző napszakokban. A felmérés célja annak feltérképezése volt, hogy az egyetem tereit és épületeit mennyire használják a polgárok. A Smart CEI Moncloa egy olyan funkciókészletet mutatott be, amely egy nagyszerű gyakorlati kísérletet jelentett a Smart City szolgáltatások számára. Az egyetem meglehetősen nagy területet foglal el (5,5 km^2), amely integrálódott a spanyol fővárosba. A nagyvárossal szimbiózisban létező egyetemi terület biztosítja a kísérlet számára az emberek és a járművek nagy forgalmát. A területen nagyszámú érzékelő (77) található az UPM különböző épületeiben. Ezek az érzékelők a két legszélesebb körben használt hardverplatformra támaszkodnak, ezek: a Raspberry Pi és az Arduino. A platformok is különösen releváns protokollokat használnak az IoT-hez, például ilyen az MQTT és egy, az RWD-re épülő, felhasználóbarát webes felület, amely lehetővé teszi bármely rendelkezésre álló eszköz (például okostelefon, táblagép, laptop) elérhetőségét. Ezenkívül a két kezdeti kísérleti szolgáltatás (nevezetesen az emberek áramlásának nyomon követése és a környezeti monitoring) szintén különösen fontos a Smart City forgatókönyvek modellezésében. A mérések alapján a következő felhasználási eseteket tartották a leghasznosabbnak.

Okos vészhelyzetkezelő alkalmazás (Smart Emergency Management Application, SEMA). A közbiztonság terén az emberek áramlás-ellenőrzésének egyik fő alkalmazási területe a SEMA. A hálózathoz csatlakozók száma alapján létrehozott hőtérképek lehetővé teszik azon területek azonosítását, ahol nagyobb az emberek koncentrációja. Így egy jövőbeli alkalmazás összehasonlíthatja az érzékelt eszközök számát egy adott küszöbértékkel a túlzott kapacitás kimutatása és elkerülése érdekében. Ez a szolgáltatás olyankor lehet hasznos, amikor valamilyen vészhelyzetben (például katasztrófák, tűz vagy földrengés) az épületekben tartózkodóknak létfontosságú, hogy melyik terület biztonságos, azaz hova menekülhetnek.



4. ábra

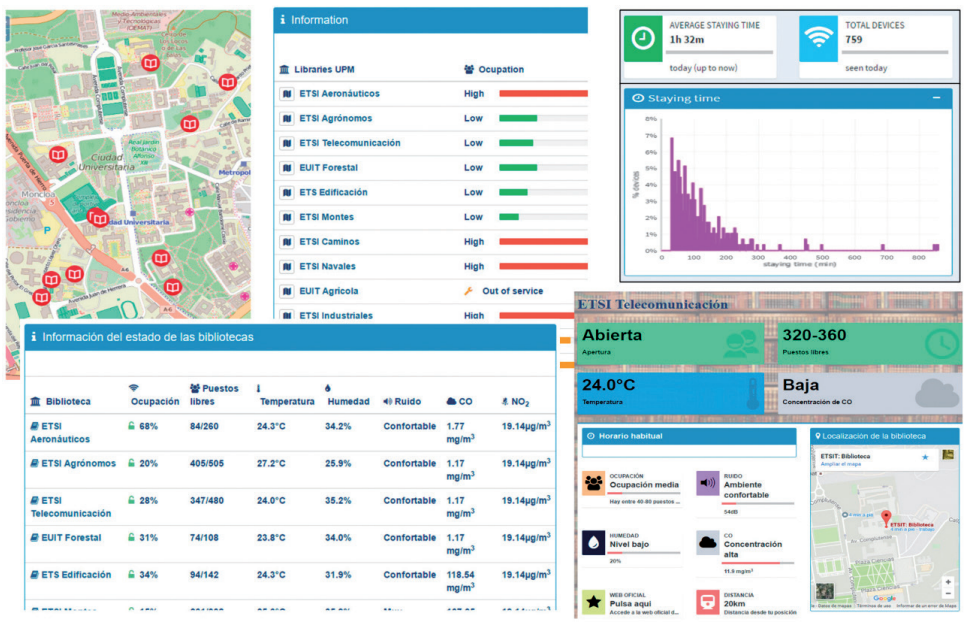
Okos vészhelyzetkezelő alkalmazás: egy épület hőtérképe

Forrás: Alvarez-Campana et al. 2017

Az okos vészhelyzetkezelő alkalmazás a smart CEI Moncloa emberáramlás-mérő szenzorhálózatán keresztül gyűjtött adatok alapján képes volt meghatározni, hogy az emberek hol találhatók. Az alkalmazás hatékonyságának vizsgálatához az adatokat, az iskolában tartózkodó diákokat, illetve azok csatlakozott eszközeinek számát vették alapul, amelyeket

decemberben gyűjtöttek. A kutatás során olyan alkalmazást fejlesztettek ki, amely feldolgozta a wifiadatokat az eszközök helyének megbecsléséhez. Az 5. ábra bemutatja azt a grafikus felhasználói felületet, amely lehetővé teszi, hogy kövessük a készülékek eloszlását az épületen belül adott időszakban.

Könyvtári alkalmazás. Ez a felhasználási eset arra példa, hogy miként kombinálható az emberek mérhető áramlása az érzékelők és a környezeti szenzorok információi révén. Ez (különösen ebben az esetben) az egyetemi közösség számára hasznosítható alkalmazás létrehozását segítette. A 6. ábra egy UPM-hallgató által kifejlesztett könyvtáralkalmazást mutat be. Ez az alkalmazás lehetővé teszi az iskola egyik könyvtárának kiválasztását, és hasznos információkat szolgáltat arról. Az alkalmazás mutatja a kiválasztott könyvtár foglalkozásait, a szabad helyek százalékos arányát (ami a wifi által észlelt eszközökön alapuló becslés), valamint a hőmérsékletet, a zajszintet, a CO_2- és NO_2-koncentrációt (amelyek a beltéri környezeti érzékelőktől származnak). Az utóbbi információ a hallgató számára lehet érdekes, mivel ezek a paraméterek hatnak a koncentrációra, és így a felsőoktatásban töltött idő hatékonyságára is. Ez az információ kombinálható a felhasználók és a könyvtárak közötti távolság adataival is, így az optimális tanulóhely kiválasztását segítheti.[3]



5. ábra

Az UPM könyvtári alkalmazás felülete

Forrás: Alvarez-Campana et al. 2017

[3] Alvarez-Campana et al. 2017.

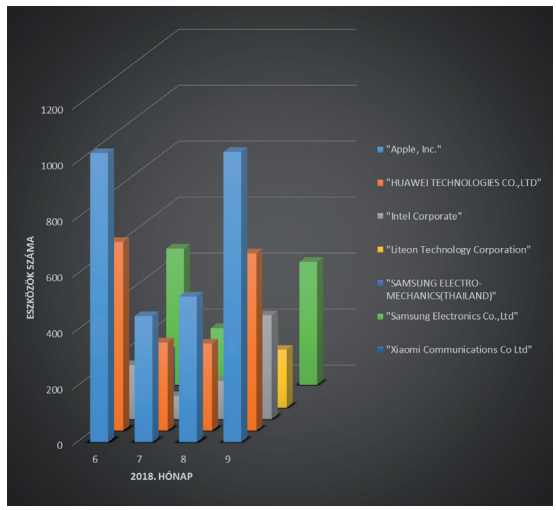
Kik használják a wifihálózatokat a Ludovika Campuson?

Az egyetemi polgároknak, akik rendelkeznek Neptun-azonosítóval, és azoknak, akik más munkát végeznek az egyetemen, hozzáférésük van az egyetemi hálózatok valamelyikéhez. Ezenkívül az egyetemre érkező vendégek számára is biztosított az egyetem hálózatain keresztüli internetelérés. A campus területén öt hálózatot nevesít az egyetem 2018-ban érvényes informatikai útmutatója:

- Az eduroam-rendszer szolgáltatásait azok a felhasználók vehetik igénybe, akik rendelkeznek érvényes felhasználói azonosítóval és hozzá tartozó jelszóval az egyetem belső Novell-hálózatában vagy az eduroam-hálózathoz csatlakozott hazai vagy külföldi intézményben (jellemzően az állandó – tanári, dolgozói – állomány tagjai).
- NKE-D: A hallgatóiwifi-szolgáltatás felhasználóazonosítás után vehető igénybe, és szabályozott sávszélességű internetelérést biztosít a hallgatóknak.
- NKE-T: Az állandó állomány számára biztosított (oktatói, tanári, dolgozói) wifiszolgáltatás is felhasználóazonosítás után vehető igénybe. Ez szabályozott, a hallgatókétól elkülönített, szabad internetelérést biztosít a felhasználóknak. Igénybevételéhez érvényes Novell-azonosítóra és jelszóra van szükség.
- NKE-GUEST: A konferenciák, rendezvények idejére a résztvevők számára biztosított biztonságos wifiszolgáltatás, amely a rendezvény résztvevői számára ismertetett, közös belépési név és az adott időszakban érvényes jelszó megadása után biztosít szabályozott sávszélességű internetelérést. Az NKE-GUEST-azonosító ellenére ez nem egy általános, nyílt elérést biztosító rendszer. A konferenciák, rendezvények idejére a résztvevők számára biztosított wifiszolgáltatáshoz történő kapcsolódáskor egy úgynevezett Captive portal jelentkezik be, amely a rendezvény résztvevői számára ismertetett, közös belépési név és az adott időszakban érvényes jelszó megadása után biztosít szabályozott sávszélességű internetelérést.[4]

A 6. ábra az egyetemi wifihálózatot használók eloszlását mutatja, ahol a felhasználók számát havi bontásban, illetve az általuk használt eszközök szerint csoportosítva ábrázoltuk. A nyári időszakban a központi router által rögzített adatok alapján Apple, Intel és Samsung által gyártott készülékeket használtak az egyetemi polgárok és dolgozók. Valószínűsíthető, hogy a 7. és a 8. havi felhasználók főleg az egyetem dolgozói közül kerültek ki. Ezekből az adatokból az is kiolvasható, hogy átlagosan nincs nagy különbség a dolgozók és más egyetemi polgárok eszközválasztása között.

[4] https://servicedesk.uni-nke.hu/



6. ábra

A Ludovika Campus négyhavi wififorgalma eszközökre bontva

Forrás: a szerzők szerkesztése az NKE naplóadatai alapján

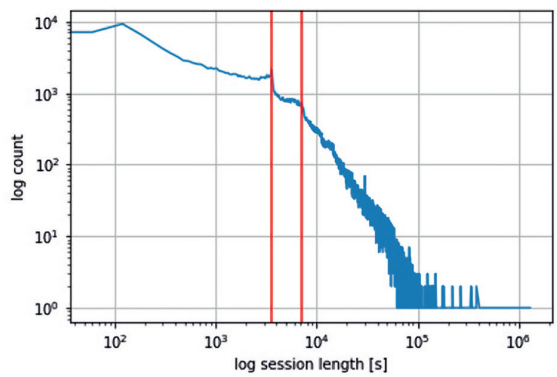
Általános statisztikai jellemzők

A kampusz központilag konfigurált és szolgáltatott wifihálózatának forgalmi adataival dolgoztunk. A kampusz különböző élethelyzeteinek megismerése érdekében egy átfogó, négy hónapos időszak bejegyzéseit elemeztük. Az időszak június 1-től szeptember 30-ig tartott, így lehetőségünk volt adatot gyűjteni egy vizsgaidőszak, egy nyári szünet és egy aktív szorgalmi időszak mintázataiból. Ezen időszak alatt több mint 2,5 millió eseménybejegyzés történt, ami a wifi-hozzákapcsolódási, wifilekapcsolódási és a wifikapcsolat-ellenőrzési bejegyzésekből tevődött össze. Ezt a forgalmat összesen valamivel több mint 5600 eszközzel bonyolították le. Tekintve, hogy a kampuszon kb. 5300 felhasználó személy fordult meg és jelentkezett fel a wifihálózatra, megállapíthatjuk, hogy egy személy átlagosan egy eszközt használt a kapcsolódáshoz. Ennek megfelelően egy eszköz az esetek többségében egy személyhez lenne kapcsolható, ezért az adatokat feldolgozás előtt anonimizáltuk: a MAC-címek utolsó három jegyét véletlenszerű kóddal helyettesítettük. A véletlenszerűsített kód esetleg felvehetett olyan értéket, ami megegyezett egy másik MAC-címmel. Mivel a statisztikai kiértékelésben nem játszik szerepet egy-egy MAC-cím egyedi története és különösen a felhasználóhoz való társíthatóság, ezért az anonimizálás nem jelentett problémát az elemzés számára. Az elemzés során felhasználónevet, Neptun-kódot, illetve más, külső forrásból személyhez köthető adatot nem használtunk fel, ezért az adatok teljes mértékben anonimnak tekinthetők.

Az adatok jellemzéséhez fontos megjegyezni, hogy a wifikapcsolódásokat egy központi azonosítórendszer szabályozza. Elemzésünkben e kampuszszintű rendszer feljegyzéseit elemezzük, tehát e szerint vizsgáljuk a kampuszon belüli mozgást.

A wifikapcsolatok időbeli főbb jellemzői

Amikor egy eszköz a wifihálózathoz csatlakozik, akkor egy előre meghatározott protokoll szerint létrejön egy kapcsolat. Ezt a kapcsolatot a rendszer időnként ellenőrzi, illetve megszakítja. A megszakításnak lehet oka, hogy a felhasználó kilépett a wifi hatóköréből, kikapcsolta az eszközét, vagy egyéb technikai probléma is okozhatja a kapcsolat bomlását. A wififorgalmat tehát alapvetően meghatározzák ezek a rádiós kapcsolatok. A vizsgált négy hónap alatt összesen valamivel kevesebb mint 260 000 kapcsolódás történt. Ezek között vannak hosszabbak, és vannak rövidebbek is.



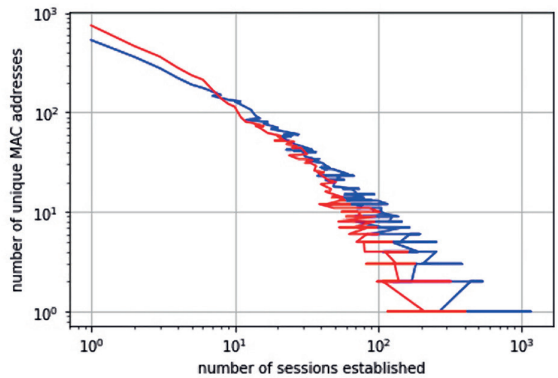
7. ábra

A kapcsolatok hosszának eloszlása a kampusz wififorgalmában (log-log ábrázolás)

Forrás: a szerzők szerkesztése

Alapvetően két viselkedési típust különböztethetünk meg a wifikapcsolatok hossza alapján: rövid kapcsolatok, amelyek legfeljebb egy óra hosszat tartanak, illetve két óránál hosszabb kapcsolatok. A rövid és hosszú kapcsolatok tartományát egy-egy kiemelkedően gyakori, percre pontosan azonos hosszúságú kapcsolódási időtartam választja el. Ez a két csúcs az óra közti szünetek, illetve a tipikusan kétórás egységekben tartott előadások ütemével magyarázható. Néhány eszköz esetén extrém hosszú kapcsolódási időt figyelhetünk meg.

A felhasználói viselkedés másik fontos kérdése az, hogy milyen gyakran kapcsolódnak fel a felhasználók a wifihálózatra. Az alábbi ábra egy tipikus emberi viselkedési görbét mutat, amely egy lassan lecsengő eloszlásfüggvénynek felel meg.



8. ábra

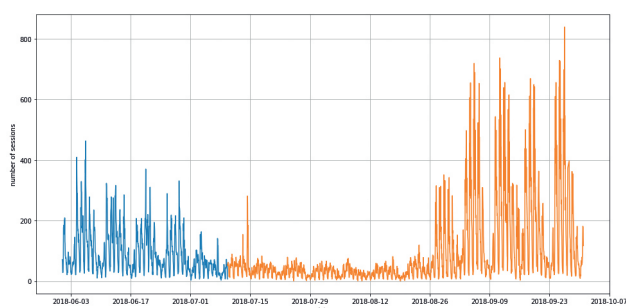
Egy eszköz által létrehozott kapcsolatok száma

Megjegyzés: A piros görbe a hosszú, legalább egyórás kapcsolatok eloszlását mutatja, míg a kék görbe a tetszőlegesen rövid kapcsolatokat is számba veszi.

Forrás: a szerzők szerkesztése

A 8. ábrán egy log-log grafikonon lecsengő görbéket látunk. Ez azoknak az eszközöknek a száma, amelyeket kevés alkalommal használtak (például maximum 10-szer) – az ilyenek igen sokan vannak. Ugyanakkor előfordulnak olyanok is, amelyek kiemelkedően sokszor kapcsolódtak a hálózathoz: néhányan több ezerszer is le- és felkapcsolódtak a wifire. A kevés kapcsolatot kezdeményező eszközök (és a fentiek szerint a kevés alkalommal csatlakozó felhasználók) valószínűleg egyedi látogatók vagy egyébként másként csatlakozó felhasználók lehetnek. Az extrém sok kapcsolódás mögött valószínűleg technikai hiba állhat, ami folyamatosan ledobja az eszközt, de az automatikusan újracsatlakozik.

A statisztikai jellemzők áttekintése után nézzük meg a kapcsolódások időbeli alakulását! A felhasználói viselkedés időbeli jellemzőit legjobban a következő ábráról olvashatjuk le.



9. ábra

Óránkénti kapcsolatfelvételek száma az idő függvényében

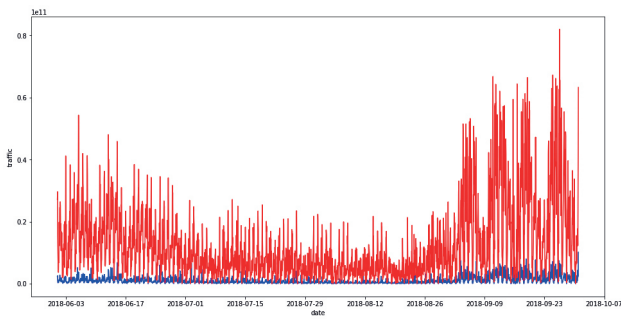
Megjegyzés: A két különböző szín a wifikapcsolatokat fogadó két központi eszközt jelöli.

Forrás: a szerzők szerkesztése

A 9. ábrán a központi logfájlban található két központi egység bejegyzéseit láthatjuk kék és sárga színnel. A vízszintes tengelyen a vizsgálat tárgyát képező négyhavi idő, a függőleges tengelyen pedig az óránként létrehozott új kapcsolatok számát láthatjuk. A két szín egymáshoz kapcsolódásából látható, hogy a diplomaosztóra készülve kicserélték a régi eszközt egy újra.

További érdekesség, hogy jól megfigyelhető a három tipikus életszakasz a kampuszon: közepes aktivitás a vizsgaidőszak munkanapjain, erős aktivitás a szorgalmi időszak első hónapjában, de csak a hét három-négy munkanapján, és végül a csendes időszak a nyári vakáció idején. Az aktivitási csúcsok változásában megfigyelhetünk egy heti ciklust végig a négy hónap alatt, azaz hétköznapokon még a nyári szabadságolások idején is nagyobb az aktivitás, mint a hétvégéken. Érdekes azt is látni, hogy a szorgalmi időszak idején a hétvégi aktivitás a nyári hétköznapok aktivitásának szintjén mozog. Az időbeli viselkedés utolsó szempontja, amelyet leolvashatunk az ábráról, a napon belüli ciklus: éjszaka kisebb az aktivitás, nappal pedig nagyobb egy déli időpont körüli maximummal.

A felhasználói viselkedés másik szempontja lehet a letöltött és feltöltött adatok mennyiségének változása. Összevetve a 9. ábra kapcsolódási számaival, érdekes viselkedési formákat ismerhetünk fel.



10. ábra

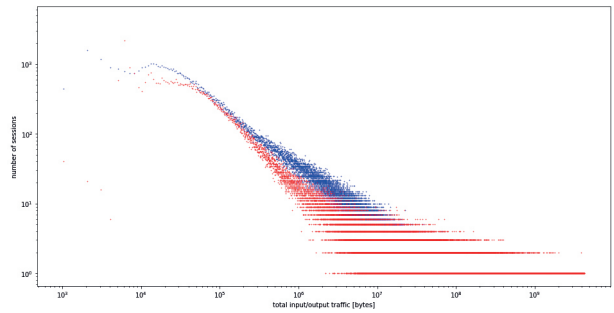
Óránkénti adatforgalom 100 Gb egységben

Megjegyzés: piros színnel a letöltött, kék színnel a feltöltött adatmennyiség látható

Forrás: a szerzők szerkesztése

A 10. ábrán a teljes kampusz wifihálózatának adatforgalmi mennyisége látható órás egységekre bontva. Piros színnel a letöltések, kék színnel a feltöltések láthatók. Természetesen a letöltések egy nagyságrenddel nagyobb mennyiséget mutatnak, mint a feltöltések. A négy hónapot átfogó adatsorban az adatforgalom hasonlóan ciklusos, többperiódusú viselkedést mutat, mint a kapcsolódások száma. Érdekes eltérés azonban, hogy a nyári időszak és a hétvégi időszakok nem jelennek meg olyan erős minimumokként, sőt, a szorgalmi időszak elején az egy-egy hétvégén letöltött adatmennyiség a hétköznapi forgalommal összemérhető nagyságrendbe esik.

Az általános statisztikák bemutatását az adatforgalom gyakoriságeloszlásával zárjuk. Az alábbi ábrán a letöltött (piros) és feltöltött (kék) adatmennyiség látható.



11. ábra

Az egy kapcsolaton belül feltöltött és letöltött adatmennyiség gyakorisága

Megjegyzés: a kék szín jelöli a letöltést, a piros a feltöltést.

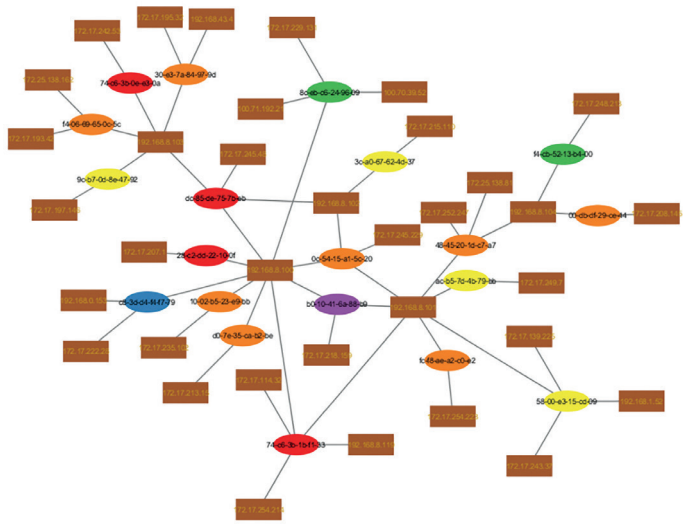
Forrás: a szerzők szerkesztése

Az eloszlás elején látható trend valószínűleg az e-mail- vagy az operációs rendszerek kisforgalmú adatkapcsolatait jelzi. A nagyobb adatmennyiségek pedig a tipikus hatványfüggvény szerinti lassú lecsengéssel jelzik, hogy vannak igen nagy letöltések is, de ezek ritkák.

Hálózatok wifibiztonsági elemzésekhez

A bevezető, általános viselkedési jellemzők után hálózattudományi eszközökkel is elemezzük az adatsort. Elsőként azt vizsgáljuk, hogy egy kapcsolat során a wifin dolgozó felhasználó milyen stabil forgalmat láthat, illetve a belső hálózaton egy eszköz milyen hozzáférési bejegyzéseket generálhat.

A wifihálózatra csatlakozott eszköznek a forgalom szempontjából két fontos azonosítója van. Az egyik az úgynevezett MAC-cím, amely a hálózat eléréséhez használt eszközt azonosítja. A másik az úgynevezett IP-cím, amely a kampusz belső hálózatához való csatlakozást biztosítja. A MAC-cím többnyire állandó, az eszközhöz kötött, az IP-cím dinamikusan változó, kapcsolódásonként más és más lehet, tehát az internet számára úgy látszik, mintha megszűnne a régi, és keletkezne egy új eszköz.



12. ábra

Az egy wifikapcsolaton belül használt IP-címek kapcsolata az eszközazonosító MAC-címekkel

Forrás: a szerzők szerkesztése

A forgalmi adatok vizsgálatában több olyan esetet is azonosítottunk, amikor a wifikapcsolat folyamatos volt, mégis a felhasználó gépe menet közben új IP-címet kapott. Az ilyen váltások a gyakorlatban azt jelentik, hogy ha egy hosszabb fájlletöltés vagy távoli bejelentkezés közben változik meg az IP-cím, akkor az adatfolyam megszakadhat, a távoli elérés „lefagy". Az interneten ugyanis egy távoli szerver vagy a kampuszon működő gép az IP-cím szerint küldi az adatokat. Az egyszerű adatvesztésen kívül még nagyobb problémát okozhat az IP-cím megváltozása, ha a korábban használt IP-címet hirtelen egy másik gép kapja meg. Egy rosszindulatú felhasználó és egy titkosítatlan adatkapcsolat komoly problémákat okozhat. Az alábbi hálózatos ábra az egy kapcsolaton belül használt IP-címeket köti össze az eszköz azonosítójával. Az eszközöket ovális csúcsok jelzik, az IP-címeket barna téglalapok.

A 12. ábra mutatja azokat az IP-címeket is, amelyeket több eszköz is használt. További tájékoztatásként azt is feltüntettük, hogy a közös IP-címet használó eszközök milyen más IP-címeket használtak még. A félreértés elkerülése érdekében megjegyezzük, hogy egy adott pillanatban egy IP-cím csak egy eszköznek van kiosztva! Tehát az ábrán látható „közös használat" csakis időben egymást követő szakaszokon valósulhat meg, a párhuzamos használat kizárt.

Látható, hogy az IP-címek jelentős része kizárólag egyetlen eszközhöz kapcsolódik, fokszáma 1. Azonban vannak olyan IP-címek, amelyek viszonylag magas fokszámúak a hálózatban, azaz több eszköz is használta őket.

Az IP-címeket váltó wifikapcsolatok idősorelemzésével megvizsgáltuk, hogy az IP-cím váltása milyen felhasználói „élményt" adhatott. Az időbeli viselkedés szempontjából két fő típust különíthetünk el az IP-váltó wifikapcsolatokra.

a) Az egyik típusban átmeneti zavar következtében kap új címet az eszköz. Itt a kapcsolat folyamán lényegében egyetlen IP-címet használ az eszköz, de néhány másodpercre átmenetileg átvált egy ideiglenes címre. Ezek a váltások olyan rövidek, hogy ezt az internetes adatkapcsolati protokollok sikeresen áthidalhatják. Mégis a néhány másodperces váltás ideje alatt adatcsomagok veszhetnek el, ami a kapcsolat lelassulását vagy lassú, akadozó programviselkedést okozhat.
b) A másik típus esetén az IP-cím-váltás végleges. Ilyenkor a kezdeti IP-címről, némi átmeneti IP-cím-ugrálás után, egy másik, állandó IP-címet kap az eszköz. Ezt a váltást már nem fogják kezelni a szabványos protokollok, ilyenkor a felhasználó úgy látja, hogy a korábbi internetkapcsolatai „megfagynak", illetve leállnak. Természetesen újraindítás vagy a kapcsolat/letöltés újrakezdése esetén már nem lesz probléma. Megjegyezzük, hogy az ilyen IP-váltások esetén különösen fontos szerepet kapnak a biztonságos protokollok. Például egy webcím böngészése esetén a titkosítatlan http-kapcsolat adatai az IP-cím alapján azonosítják a gépeket. Ha egy gép címe megváltozik, és a helyébe egy másik gép kerül, akkor a megkezdett folyamatot átveheti a másik gép. A biztonságos, titkosított kapcsolatok esetén (például https-protokoll használatával) már hiába veszi át az esetleg rosszindulatú új felhasználó az adatfolyamot, a titkosítási kulcs hiányában nem tudja értelmezni a kapott információkat.

Hálózatok felhasználói viselkedés vizsgálatához

Projektünkben az adatok felbontása nem adott lehetőséget a felhasználói viselkedés olyan részletes elemzésére, mint amelyet a madridi egyetem kutatói végeztek. Esetünkben a teljes kampusz viselkedése látható, így ha kisebb csoportok, együtt mozgó csoportok azonosítását szeretnénk elvégezni, akkor a térbeli azonosítás helyett csupán az időbeli azonosítás áll rendelkezésünkre.

Első megközelítésben azt várjuk, hogy ha két felhasználó egyszerre használja a wifit, akkor valamilyen közös használati szokásban hasonlíthatnak egymásra. Azaz ha a felhasználókat azonosítjuk az általuk használt eszközzel, akkor az egy időben dolgozók hálózatát meg lehet konstruálni.

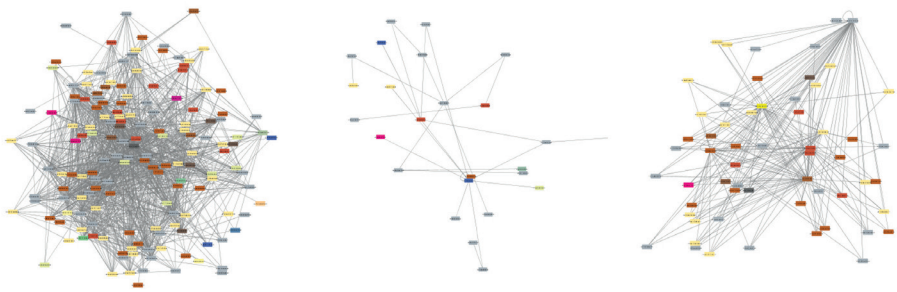
Ez a hálózat azonban nagyon extrém tulajdonságokkal rendelkezik. Egyrészt a nagyon hosszú ideig wifin maradt eszközök nagyon sok felhasználót kötnek össze, másrészt napközben, a legintenzívebben használt időszakban a wifit együtt használók hálózata több millió éllel rendelkező, 1000 × 1000 csúcsból álló, teljesen összekötött alhálózatokat definiál. Ilyen nagy sűrűségű és sok csúcsból álló hálózatok elemzése csak fejlett programozási eszközökkel lehetséges, emberi megtekintésre, vizualizálásra alkalmatlanok. További probléma a pillanatról pillanatra együtt bejelentkezettek hálózatában, hogy a nagy összekötöttség ellenére a fluktuáció jelentős.

Stabil mintázatok kereséséhez érdemes kihasználni a 9. ábrán mutatott időbeli viselkedést. Egyrészt a hálózatban várható egy heti ismétlődő ciklus, másrészt várható egy napon belüli, váltakozó felhasználói viselkedés. Erre építve definiálhatunk egy olyan hálózatot, amely csupán az időbeli korrelációk alapján kimutathatja a különböző felhasználói csoportokat. A kampusz heti rendszerességű programjain (például tanórák) kitartóan részt vevőket úgy azonosíthatjuk, hogy kiválogatjuk azokat az eszközöket, amelyeket egy hónapon keresztül

a hét azonos napjain azonos órában használnak (például minden kedden reggel 7 óra és 8 óra között). A hosszan bejelentkezve maradottak azonban ezt a képet is túlságosan összekuszálják, ezért a hálózatból célszerű őket kizárni. Ezért az együttesen, korreláltan viselkedő felhasználók csoportjainak hálózatát úgy definiáltuk, hogy csak az azonos időpontban bejelentkezőket vesszük be a hálózatba.

Az így definiált hálózatban tehát a csúcsok azok a hálózatra kapcsolódó eszközök, amelyek a hét azonos napjának azonos időpontjában a wifihálózaton vannak, a kapcsolat pedig akkor áll fenn két csúcs között, ha legalább három alkalommal ténylegesen együtt, azonos órában jelentkeznek be. Ezzel a definícióval egy olyan hálózatsorozatot kapunk, amely időben változik, de mégsem teljesen összekötött, belső struktúra nélküli csoportokból épül fel.

Az alábbi ábrasorozaton bemutatjuk egy „átlagos" nap történetének három tipikus lépését.



13. ábra

A wifire rendszeresen együtt belépők hálózata

Megjegyzés: balról jobbra: 6–7 óra között, 11–12 óra között, 18–19 óra között.

Forrás: a szerzők szerkesztése

Látható, hogy a reggel együtt ébredők, reggelizők egy tipikus csoportot alkotnak. A dél környékén belépők alkotják a legnagyobb hálózatot, míg az esti belépők hálózata már igen ritka.

A 12. ábra és a 13. ábra hálózataiban az eszközöket reprezentáló csúcsokat különböző színekkel töltöttük ki. A színekkel azt különböztetjük meg, hogy egy eszköznek mi a gyártója. Ez a metaadat további szociológiailag érdekes elemzéseket tesz lehetővé. Vizsgálható ugyanis, hogy például az Apple-eszközök felhasználói tipikusan korán kelők vagy inkább későn kelők-e. De választ kaphatunk olyan kérdésekre is, amely a társadalmi szeparációkra világíthat rá. Például az azonos árkategóriába tartozó eszközök felhasználói gyakrabban alkotnak-e csoportokat? Vagy az eltérő anyagi lehetőségekkel rendelkező felhasználók keverednek egymással, és gyakran tudnak együtt dolgozni, vagy legalábbis az internetet együtt szokták-e használni?

Összefoglalva, az internet adatforgalma folyamatosan növekszik a világban, amelyre egyre inkább a mobil- és okostelefonokat használják. Az okosmobileszközök használatának a kényelem mellett kockázatai is vannak, ha óvatlanul használjuk őket. Főleg akkor, amikor wifihálózathoz csatlakoztatjuk eszközeinket, főleg ha az a hálózat nyílt hozzáférésű.

A kockázatoknak is több aspektusa létezik. Beszélhetünk informatikai és adatvédelmi kockázatokról, de ezek együttes jelenléte sem ritka. Az informatikai biztonságot főleg tudatossággal tudjuk megteremteni. Azonban az adatvédelmi biztonság, ahogy láthattuk, nagyrészt rajtunk kívül álló tényezőktől függ, ha használjuk az ilyen eszközöket.

Az adatvédelmi szempontoknak megfelelve, a tudományos kutatások számára nagy segítség lehet az okosmobileszközök nagyszámú használata. A felhasználói szokások megfigyelésével több, eddig nehezen megoldható probléma válik megoldhatóvá. A társadalomtudományi kérdések mellett olyan kérdésekre is választ kaphatunk, most csak néhány példát kiemelve, mint a közlekedésszervezés optimalizálása, a vészhelyzeti stratégiák javítása vagy az egyetemi oktatásszervezés hatékonyabbá tétele. A fejezetben vizsgált időszakban a kampusz életébe nyerhettünk rövid betekintést. A felhasználói szokások minőségét vizsgáló kutatásunk a kiszámítható eredmények mellett produkált nem várt információkat is. Az eszközök csatlakozási hálózatán keresztül megtudhattuk, hogy milyen termékpreferenciáik vannak a csoportosan együtt mozgó felhasználóknak. Az egyetem életét vagy órarendjét tanulmányozva az is kikövetkeztethető, milyen szakra járhatnak az azonos internetezési szokásokkal rendelkezők. Ezenkívül olyan társadalmi összefüggésekre is következtethetünk ezekből az adatokból, hogy milyen társadalmi összetételűek az egyetem polgárai, van-e eltolódás preferencia- vagy anyagi viszonylatban.

Felhasznált irodalom

A Nemzeti Közszolgálati Egyetem informatikai tájékoztatója. Elérhető: www.uni-nke.hu/egyetem/szabalyzatok-dokumentumok/egyeb-szabalyzatok

Alvarez-Campana, M. – López, G. – Vázquez, E. – Villagrá, V. A. – Berrocal, J. (2017): Smart CEI Moncloa. An IoT-based Platform for People Flow and Environmental Monitoring on a Smart University Campus. *Sensors,* Vol. 17, No. 12. DOI: https://doi.org/10.3390/s17122856

Cole, T. L. – Barber, S. eds. (2007): *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications (IEEE Std 802.11-2007).* Local and Metropolitan Area Networks, Specific Requirements, IEEE Standard for Information technology – Telecommunications and information exchange between systems. Piscataway (NJ), Institute of Electrical and Electronics Engineers.

Oberhaus, D. (2017): How a Wi-Fi Pineapple Can Steal Your Data (And How to Protect Yourself From It). *Motherboard.com,* 2017. 11. 20. Elérhető: https://motherboard.vice.com/en_us/article/pa39xv/pineapple-wifi-how-to-mitm-hack (A letöltés dátuma: 2019. 03. 12.)

Vákát oldal

Sasvári Péter – Urbanovics Anna – Tibély Gergely – Palla Gergely

Társadalomtudományi doktori iskolák társpublikációs hálózatának elemzése

Bevezetés

Magyarországon a felsőoktatásban a bolognai oktatási rendszer érvényesül, tehát a legtöbb képzés 3+2 éves bontásban valósul meg. Ezt követően a hallgató doktori képzésre jelentkezhet, amelynek elsődleges feladata a kutatásban való elmélyülés segítése. Ellentétben számos nyugati ország modelljével, Magyarországon a doktori képzésekben részt vevők doktori iskolákba szerveződnek. Mára már a legkülönfélébb tudományterületeknek is van doktori iskolájuk, amelyek közös kutatási intézetként is működnek. A doktori képzés jelenleg 4 évig tart, amelyből az első 2 év kurzusokba rendeződik, ahol különböző ismereteket, mély tudományos ismereteket kapnak a hallgatók. Ezt követően a 3. és 4. évben már egyrészt a gyakorlati ismeretekre (oktatásra), valamint a doktori disszertációra és az elegendő publikációs tevékenységre (pontok alapján) helyeződik a hangsúly. A társadalomtudományi területen összesen 11 tudományág köré rendeződnek a doktori iskolák. Míg a legtöbb doktori iskola a gazdálkodás és szervezés tudományterületein található, addig vannak egészen speciális területekre épülő doktori iskolák is, amelyekből csak 1-1 létezik.

A doktori iskolák a magyarországi felsőoktatás jövőjének letéteményesei. Egyrészt gondoskodnak a tudományos utánpótlásról, felkészítik a hallgatókat a PhD-fokozat megszerzésére, másrészt az adott egyetem szellemi műhelyeként az alap- és alkalmazott kutatások ösztönzésében, kivitelezésében is részt vesznek, harmadrészt pedig koordinálják a habilitációs eljárásokat.[1] A nappali és levelező, továbbá az egyéni doktori képzésben részt vevők jelentkezésénél eltérő meggondolások játszanak szerepet. A levelező és egyéni képzés felé orientálódók többnyire egyrészt munkahelyük megtartása, illetve előmenetelük céljából, másrészt presztízsokokból vágnak neki a doktori tanulmányoknak. A nappali képzésre jelentkezők szinte kivétel nélkül a frissen végzett hallgatók közül kerülnek ki, akiknek egy része valóban a tudományos fokozat megszerzésére törekszik, másik része viszont csupán parkolópályának tekinti a doktori képzést, és amint kedvező elhelyezkedési lehetőséget talál, abbahagyja tanulmányait és a tudományos kutatást.[2]

A doktori iskolák egyfajta tudományos kutatóműhelynek is tekinthetők, ahol az oktatók és hallgatók (doktoranduszok) azonos tudományterületen, szomszédos kutatási területeken tevékenykednek. A témavezető-témavezetett kapcsolatokból számos további kapcsolat rajzolódik ki, hálózatosodás figyelhető meg közöttük. A doktori iskolában a témavezetők és a témavezetettek írhatnak társszerzőségben is, tehát az addigi kétszereplős kapcsolati szálak kibővülhetnek mind

[1] Michalkó–Zsóka 2016.

[2] Szabó–Bánszki–Ruzsányi 2002.

horizontálisan, mind vertikálisan. Előzetes kutatásunkban kimutattuk, hogy a társadalomtudomány területén a doktori iskolai tagok 70%-a ugyanabban a doktori iskolában folytatja munkáját, ahol fokozatot szerzett, ezért a doktori iskola alapító tagjaitól kezdve az idő múlásával egy egymásra épülő összefüggő láncolat alakul ki.[3]

E tanulmány célja, hogy feltérképezze a társadalomtudományi doktori iskolákon belüli és külső kapcsolati hálókat és láncolatokat, amelyek alapján a társadalomtudomány teljes magyar vonatkozása felépíthetővé válik.

Hálózati alapfogalmak

A hálózat a matematikai gráfelmélet és a számítógép-tudomány egyik alapvető fogalmára épül. Az absztrakt gráf valamilyen objektumok *(csomópontok, csúcspontok)* és a közöttük értelmezett összeköttetések *(kapcsolatok, élek)* halmaza. Tanulmányunkban a csomópontok a szerzőket, a kapcsolatok a társszerzőségi viszonyokat reprezentálják (lásd: 1. ábra). Alapvetően nem különböztetjük meg a társszerzőség szempontjából, hogy A szerző ír-e B-vel, vagy fordítva, a hálózat élei tehát irányítatlanok.

Az *élsűrűség* megadja a két véletlenszerűen kiválasztott szerző (csomópont) közötti közvetlen kapcsolat létezésének valószínűségét. Esetünkben, ha egy szerző az adott doktori iskola minden tagjával írna közleményt, akkor a sűrűség 1 lenne, ha senkivel sem, akkor 0.

Az *összefüggő komponens* olyan részgráf, amely összefüggő, azaz bármely két csúcsát út köti össze, de a hálózat többi csúcsához nem csatlakozik.

Az általunk vizsgált hálózatban egy szerző *fokszáma* a szerző és a hálózatban szereplő többi szerző közötti kapcsolatok száma. Ez mutatja meg, hogy egy szerzőnek hány társszerzője volt eddig. Az *átlagos fokszám* az összes szerző társszerzőinek átlagos számát mutatja.[4] *A legrövidebb út hossza* alatt a két szerző közötti távolságot értjük (egyik szerzőtől a másikig eljutva az érintett élek száma), amely a legkevesebb összekötő szerzőt tartalmazza. *Átmérő* alatt a szerzők között fellépő legnagyobb távolságot értjük. Az *átlagos úthossz* a szerzők közötti távolságok átlaga.

A *klaszterezettségi együttható* az adott szerző társszerzői közötti kapcsolatokat vizsgálja. Másképpen fogalmazva azt, hogy az így létrejövő háromszögekben vajon mindenki társszerzőségi viszonyban van-e egymással.

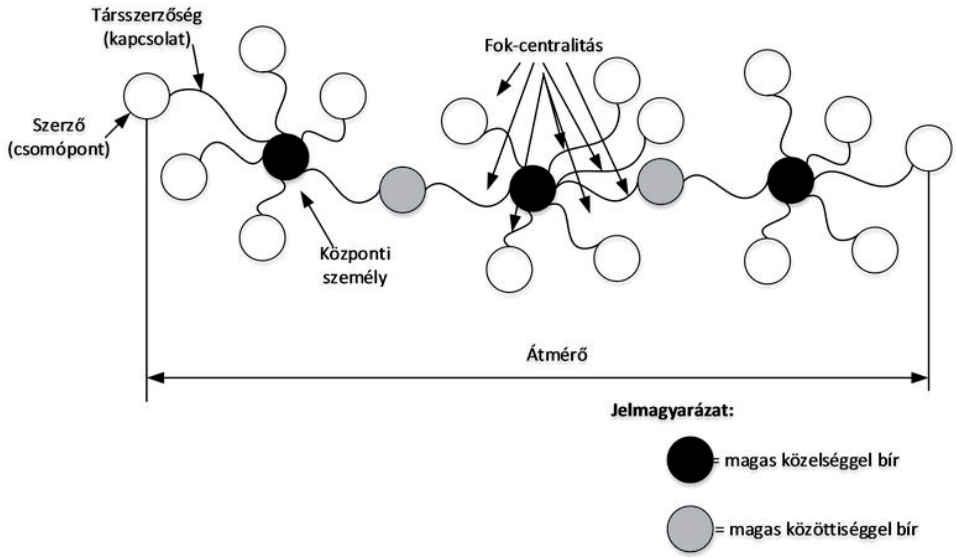
A hálózati megközelítést jól lehet alkalmazni a legfontosabb szereplő meghatározására. A fontos szereplők általában a kapcsolatháló stratégiai pontjaiban helyezkednek el, de a fontosság számítása több módon is megközelíthető attól függően, hogy mi alapján tekintünk valakit fontosnak. Tekinthetjük azt *központi személynek,* aki a legnagyobb kapcsolati aktivitást mutatja, és akivel sokan írnak, sok más szerzőnek a szerzőtársa. Az egyik jellemző centralitásszámítási mód a *fokszámcentralitás* (degree centrality, CD), ahol abból indulunk ki, hogy a szereplő aktivitását a fokszám (azaz a hozzá közvetlenül kapcsolódó más szerzők száma) jól méri.

A *közelség* (closeness centrality, CC) abból indul ki, hogy egy szereplő akkor van központi helyzetben, ha minden tagot viszonylag könnyen és gyorsan elér, így nem kell más szereplőkre hagyatkoznia.

---

[3] Sasvári 2018.

[4] Barabási 2016, 446.



1. ábra

Az alapfogalmak bemutatása

Forrás: Wasserman–Faust 1994 alapján a szerzők szerkesztése

A *köztesség* (betweenness centrality, CB) esetében a kiindulási pont az, hogy igazán azoknak a szereplőknek van jelentős befolyásolási lehetősége, akik képesek ellenőrizni a kapcsolatokon keresztül áramló erőforrásokat vagy információkat, azaz akik sok másik szereplő között helyezkednek el. Így például, ha egy adott pontból a legrövidebb út egy másik pont felé két másik szereplőn keresztül vezet, a két közbülső szereplő meghatározó lehet a kapcsolatokban (ezek a közvetítők, avagy a hídszerepű szereplők).

Vizsgálati szintek

A Magyar Tudományos Akadémia Könyvtár és Információs Központ mint akadémiai költségvetési szerv közreműködésével egy tudományos művek adatait tartalmazó nemzeti tudományos bibliográfiai adatbázist működtet, ez a Magyar Tudományos Művek Tára (MTMT). Ez a bibliográfiai hitelesség szempontjából ellenőrzött módon tartalmazza a költségvetési szerveknél foglalkoztatottak által megjelentetett tudományos műveket. (Lásd: 1994. évi XL. törvény a Magyar Tudományos Akadémiáról)

Magyarországon doktori képzés kizárólag doktori iskola keretében folytatható. A doktori iskola létesítésekor meg kell jelölni azt a tudományterületet, azon belül a tudományágat vagy művészeti ágat, amelyben a doktori képzést folytatni kívánják. A tudományágon belüli, illetve tudományágak közötti kutatási terület megnevezéssel azonosítható a doktori iskola működési kereteit tükröző szakmai tevékenység. [Lásd: 387/2012. (XII. 19.) Korm. rendelet a doktori iskolákról, a doktori eljárások rendjéről és a habilitációról]

Az elemzésben felhasznált adatok az Országos Doktori Tanács (ODT) hivatalos oldalairól származó névsorból és a szerzőnkénti tudományterületi besorolásból származnak, valamint a szerzőkhöz kapcsolt egyedi MTMT-azonosítók együtteséből állnak.

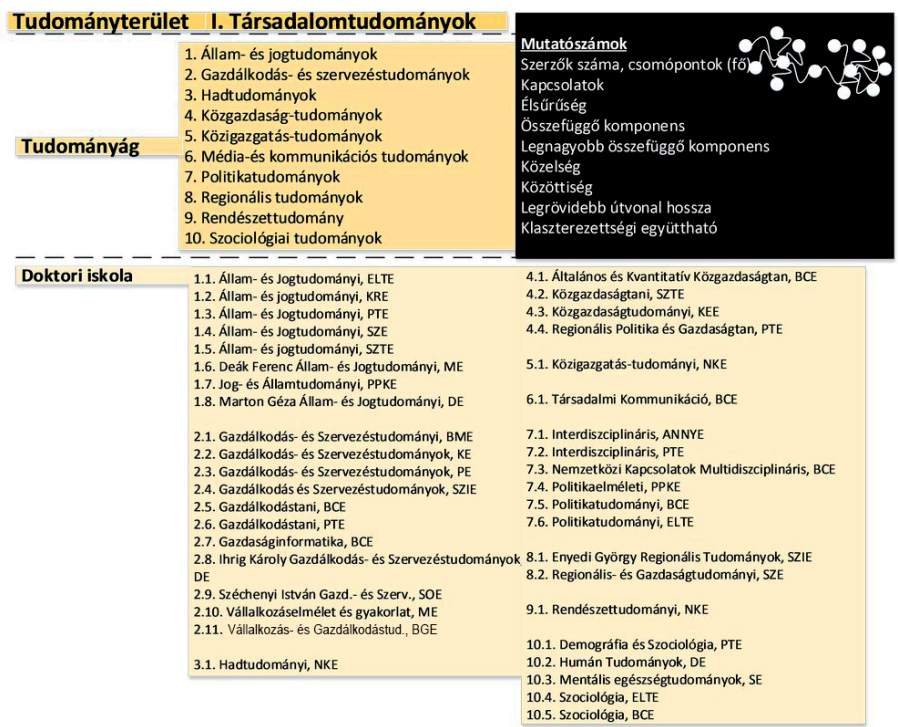
A vizsgálatunknak 3 szintje van (lásd: 2. ábra):

I. tudományterületek szintje
II. tudományágak szintje
III. doktori iskolák szintje

Az ODT weboldala alapján az alábbi tudományterületeket különböztetjük meg:

1. agrártudományok
2. bölcsészettudományok
3. hittudományok
4. műszaki tudományok
5. művészetek
6. orvos- és egészségtudományok
7. társadalomtudományok és
8. természettudományok

Jelen közleményünk vizsgálati kerete a társadalomtudományokra szorítkozik.



2. ábra

A vizsgálat szintjei és a kapcsolódó mutatószámok

Forrás: a szerzők szerkesztése

A társadalomtudományt az ODT szerint az alábbi tudományágakra bonthatjuk:

1. állam- és jogtudományok
2. gazdálkodás- és szervezéstudományok
3. hadtudományok
4. közgazdaság-tudományok
5. közigazgatás-tudományok
6. média- és kommunikációs tudományok
7. politikatudományok
8. regionális tudományok
9. rendészettudomány
10. szociológiai tudományok

A különböző társadalomtudományi tudományágakat 18 egyetemen, 39 doktori iskolában (a továbbiakban: DI) lehet tanulni. A 39 DI-ból 3 több társadalomtudományi tudományágban is kiállít oklevelet. A legtöbb DI a *gazdálkodás- és szervezéstudományok* (12 darab) és *állam- és jogtudományok* (9 darab) tudományágban működik. Ezzel szemben a *had-,* a *közigazgatás-,* a *média- és kommunikációs,* illetve a *rendészettudományok* területén csak egy, a *regionális tudományok* területén 2 DI található. A *közgazdaság-tudományok* és *politológia* területén 6 doktori iskola, a *szociológiai tudományokén* 5 doktori iskola található.

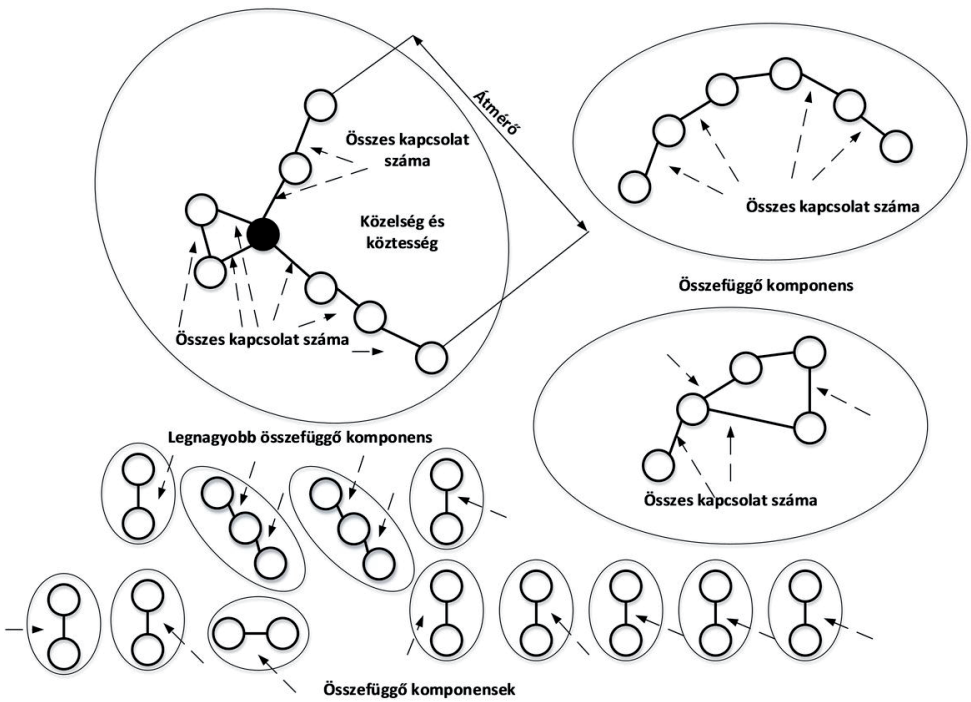
Következtetések és megállapítások

A társadalomtudományi doktori iskolák és tagjaik adatai szintén az ODT adatbázisából származnak. A tudományágak adják a tudományági felosztást az elemzés során is, mivel mindegyik doktori iskola valamelyik tudományágba van besorolva. Az elemzésben így tehát összesen 39 doktori iskolát mutatunk be, amelyek profilját egyenként vettük górcső alá. Az oktatók, témavezetők és témakiírók listáját minden doktori iskola esetén feltüntették a honlapon, ahol különböző azonosítók is rendelkezésre állnak, mint például a védés dátuma vagy a még folyamatban lévő doktori képzési folyamat státusza. Az elemzésbe ilyen szempontok szerint összesen 2148 oktató, valamint a hallgatókkal kiegészítve 3933 fő került bevonásra. Az MTMT egyedi azonosítóit az ODT-n elérhető profilból nyertük ki. Összesen 2290 darab MTMT-azonosítót töltöttünk le, majd az ezekhez tartozó publikációs adatokat is kinyertük az MTMT rendszeréből. A társszerzőségi adatokat is az MTMT egyedi azonosítói szerint szereztük be.

A társadalomtudomány jellemzői

Magyarországon 173 darab DI működik. A DI-k 21%-a a társadalom-, ötöde a bölcsészet-, 18%-a a természet-, 15%-a az orvos-, 14%-a a műszaki-, 6%-a az agrár-, 4%-a a hittudományok és 4%-a a művészetek területén tevékenykedik. Az elemzésbe összesen 2304 főt vontunk be, akik közül 481 szerzőnek nincs MTMT-azonosítója, vagy nem rendelkezik társszerzővel a saját doktori iskolájában az adatbázis tagjainak körében. Ebből adódóan 1823 fővel tudtunk társszerzőségi adatokat számolni. Ezek közül 1786 fő alkotja

a legnagyobb összefüggő komponenst. A vizsgálatunk további eredményeként 6 és 5 fős komponensből 1-et, 3 fősből 2-t és 2 fősből 10 darabot találtunk még.



3. ábra

Illusztrációs ábra a hálózattudomány területén

Forrás: a szerzők szerkesztése

A társadalomtudomány területén a kapcsolatok, a társszerzők száma meghaladja a 6 ezer főt (6069) (lásd: 3. ábra). Az élsűrűség 0,0036, az átlagos fokszám 6,65, az átmérő pedig 14 volt. A hálózat átlagos közelségi mutatója 0,195, köztességi mutatója pedig 0,0024. A társadalomtudomány *központi személye* – mind közelség, mind köztesség tekintetében – Rechnitzer János (lásd: 1. táblázat). Megjegyezzük azonban, hogy a pontos sorrend érzékeny lehet az adatok teljességére, azaz ha véletlenül kimarad az MTMT-ből egy-egy tanulmány, az befolyásolhatja a pontos sorrendet. Egy ilyen megbízhatóságelemzés azonban túlmutat e tanulmány keretein.

1. táblázat

A legnagyobb köztességgel és közelséggel rendelkező szerzők a társadalomtudományban a társpublikációk területén

Sorszám	Név	Köztesség	Sorszám	Név	Közelség
1.	Rechnitzer János	**0,0566**	1.	Rechnitzer János	**0,2707**
2.	Poór József	**0,0431**	2.	Poór József	**0,2610**
3.	Nemeslaki András	**0,0426**	3.	Czakó Erzsébet Hajnalka	**0,2553**
4.	Róbert Péter	**0,0361**	4.	Nemeslaki András	**0,2526**
5.	Fónai Mihály Ferenc	**0,0251**	5.	Lengyel Imre	**0,2526**
6.	Padányi József	**0,0242**	6.	Róbert Péter	**0,2523**
7.	Józsa László	**0,0242**	7.	Szintay István	**0,2515**
8.	Lengyel Imre	**0,0236**	8.	Kocziszky György	**0,2505**
9.	Fazekas Judit	**0,0231**	9.	Villányi László	**0,2501**
10.	Pusztai Gabriella	**0,0225**	10.	Michalkó Gábor	**0,2497**

Forrás: a szerzők szerkesztése

A tudományágak jellemzői

A legtöbb oktatóval – aki lehet törzstag, témakiíró és témavezető is – a *gazdálkodás- és szervezéstudományok* területén találkozhatunk (807 fő) (lásd: 2. táblázat). Ez talán nem is meglepő, mivel ebben a tudományágban 12 darab DI tevékenykedik. A legkevesebb tagot a *média- és kommunikációs tudományok* területén találtuk.

A vizsgált oktatók esetén elmondható, hogy az ODT-szabályozás miatt egy oktató csak egy doktori iskolában lehet törzstag, illetve egy témakiíró vagy témavezető több tudományterületen, tudományágban és DI-ben is előfordulhat.

Az átfedések miatt az ODT adatbázisában 2017-ben 2805 oktatót tüntettek fel valamely társadalomtudományi DI tagjaként.

2. táblázat

A magyarországi doktori iskolák oktatóinak száma tudományáganként 2017-ben

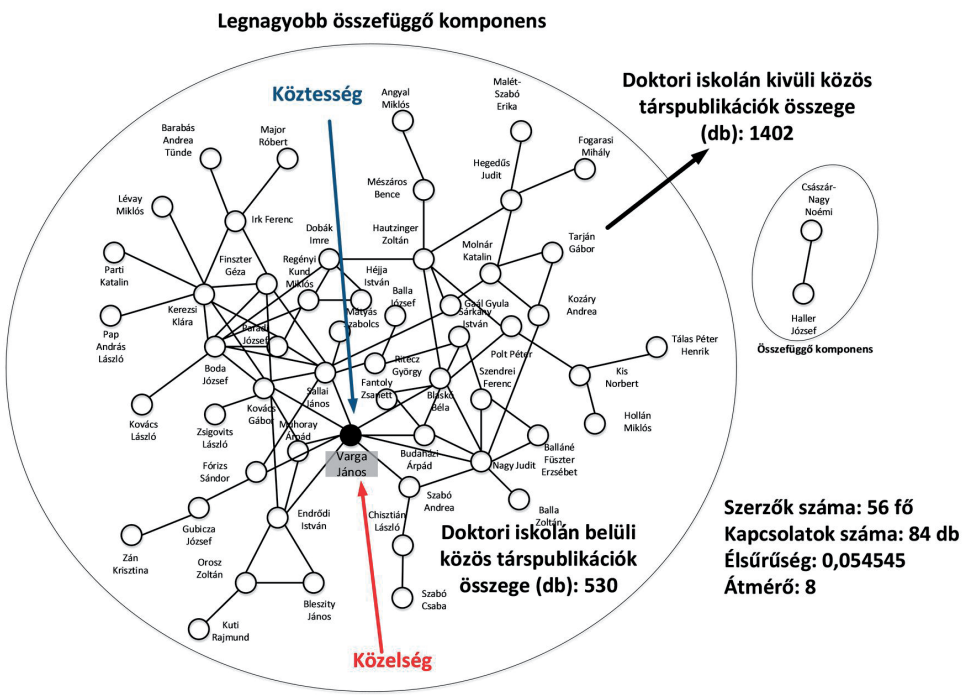
Tudományág	Doktori iskolák száma (darab)	Oktatók száma (fő)	Szerzők (csomópontok) száma (fő)	Vizsgálatban szereplők aránya (%)
Állam- és jogtudományok	9	509	371	73%
Gazdálkodás- és szervezéstudományok	11	807	693	86%
Hadtudományok	1	122	106	87%
Közgazdaság-tudományok	6	285	224	79%
Közigazgatás-tudományok	1	110	70	64%
Média- és kommunikációs tudományok	1	53	26	49%

Tudományág	Doktori iskolák száma (darab)	Oktatók száma (fő)	Szerzők (csomópontok) száma (fő)	Vizsgálatban szereplők aránya (%)
Politikatudományok	6	285	211	74%
Regionális tudományok	2	144	125	87%
Rendészettudományok	1	73	56	77%
Szociológiai tudományok	5	417	291	70%
Összesen	**43**	**2 805**	**2173**	**77%**

Forrás: az Országos Doktori Tanács adatbázisa alapján a szerzők szerkesztése

Különböző feldolgozottsági szint mellett 2173 fő szerepel az elemzésünkben csomópontként. Arányaikat tekintve az elemzésben a *regionális tudományok* képviselői közül sikerült a legtöbb oktatót (87%) vizsgálatunkba bevonni, míg a legkevesebbet a *média- és kommunikációs tudományok* (49%) területéről.

A legtöbb oktatót tömörítő tudományágak sorrendben: a *gazdálkodás- és szervezéstudományok* (693 fő), *állam- és jogtudományok* (371 fő), valamint a *szociológiai tudományok* (291 fő). A legkevesebb tag a *média- és kommunikációs tudományok* (26 fő), a *rendészettudomány* (56 fő) (lásd: 4. ábra) és a *közigazgatás-tudományok* (70 fő) területén található.



4. ábra

A rendészettudomány társpublikációs hálója

Forrás: az Országos Doktori Tanács adatai és az MTMT adatbázisa alapján a szerzők szerkesztése

A *kapcsolatok* száma is ezen eredményre rímel, mert a legtöbb társpublikációs kapcsolat a *gazdálkodás- és szervezéstudományok* (2710 darab), az *állam- és jogtudományok* (1000 darab) és a *szociológiatudományok* (639 darab) területén alakult ki, míg a legkevesebb társpublikációval a *média- és kommunikációs tudományok* (24 darab), a *rendészettudomány* (84 darab) és a *közigazgatás-tudományok* (99 darab) területén találkozhatunk (lásd: 3. táblázat). Ez az eredmény természetesen következik az adott tudományágban tevékenykedő tagok számából, tehát az általuk alkotott hálózat nagyságából is.

3. táblázat

Az egyes tudományágak jellegzetes tulajdonságai

Tudományág	Szerzők száma (fő)	Legnagyobb összefüggő komponens (fő)	Kapcsolatok, társszerzők száma (fő)	Élsűrűség	Klaszterezettségi együttható	Átlagos fokszám
állam- és jogtudományok	371	335	1000	0,015	0,329	5,4
gazdálkodás- és szervezéstudományok	693	686	2710	0,011	0,378	7,8
hadtudományok	106	95	238	0,043	0,400	4,5
közgazdaság-tudományok	224	215	487	0,019	0,336	4,3
közigazgatás-tudományok	70	63	99	0,041	0,239	2,8
média- és kommunikációs tudományok	26	10	24	0,074	0,290	1,8
politikatudományok	211	179	307	0,014	0,272	2,9
regionális tudományok	125	114	301	0,039	0,314	4,8
rendészettudomány	56	54	84	0,055	0,246	3,0
szociológiai tudományok	291	269	639	0,015	0,300	4,4

Forrás: az Országos Doktori Tanács adatai és az MTMT adatbázisa alapján a szerzők szerkesztése

Ennél pontosabb képet alkot az, ha hozzátesszük, hogy mely tudományág esetében mekkora a *legnagyobb összefüggő komponens* mérete. Ilyen szempontból azt láthatjuk, hogy a legösszetartóbb csapatot a gazdálkodás- és szervezéstudományok tudhatja magáénak (99%, 686 fő), majd őket követi a rendészettudomány 97%-kal (54 fő). A legkevésbé összetartók a *média- és kommunikációs tudományok* képviselői 39% (10 fő), de általában a 90% feletti eredmények a meghatározók. Ez azt jelenti, hogy a legnagyobb összefüggő komponens létszáma a teljes létszám több mint 90%-át teszi ki. Ez az arány igen szép eredmény, amiből arra következtethetünk, hogy a tudományágak képviselői jól ismerik egymást, és aktív társszerzői kapcsolatokat ápolnak egymással.

Az *élsűrűség,* amely az összes potenciálisan fellelhető kapcsolatok számához viszonyítja a valójában is létrejött kapcsolatok számát, ennél árnyaltabb képet ad a tudományági hálózatok összetartásáról, a társszerzői együttműködéseinek gyakoriságáról. Kitűnik, hogy a kisebb létszámmal rendelkező tudományágak összetartóbbak: *média- és kommunikációs tudományok* (0,074), *rendészettudomány* (0,055), *hadtudomány* (0,043), míg a *gazdálkodás- és szervezéstudományok* a legszétszéledőbb (0,011), ezt követik a *politikatudományok* (0,013), valamint az *állam- és jogtudományok* (0,015) oktatói.

A *klaszterezettségi együttható* tekintetében, amely arra keresi a választ, hogy mekkora az esély egy szerző két társszerzőjének társszerzőségi viszonyára, a 0,25–0,35-ös arány a legmeghatározóbb. Ez alól a *hadtudomány* jelent kivételt 40%-os (0,400) valószínűséggel, valamint a *gazdálkodás- és szervezéstudományok* 38%-kal (0,378). A legkevésbé jelentős együtthatói értékekkel a *közigazgatás-tudományok* csoportja rendelkezik (0,239), majd ezt követi a *média- és kommunikációs tudományok* és a *rendészettudomány* (0,290, illetve 0,246). Míg a legjelentősebb klaszterezettségi együtthatóval rendelkező tudományágak képviselői centrikusan helyezkednek el, jól ismerve egymást és aktívan alakítva közös kutatói műhelyüket, addig a legkisebb értékekkel rendelkezők inkább lineárisan helyezkednek el egymáshoz képest, egy-egy neves képviselőre felfűződve.

Az *átlagos fokszám* megmutatja, hogy az adott tudományág tagjai átlagosan hány társszerzővel rendelkeznek a hálózatukon belül. Ebből kitűnik, hogy a legaktívabbak a *gazdálkodás- és szervezéstudományok* (7,8 kapcsolat/fő) és az *állam- és jogtudományok* (5,4 kapcsolat/fő), míg a legkevesebb társszerzővel átlagosan a *média- és kommunikációs tudományok* képviselői (1,85 kapcsolat/fő) rendelkeznek. Ez a tudományág nagyságára is visszavezethető, mert míg a gazdálkodás- és szervezéstudományok, valamint az állam- és jogtudományok képviselői saját területükön is válogathatnak társszerzői partnereket, addig a média- és kommunikációs tudományok kutatói ilyen szempontból kevesebb mozgástérrel rendelkeznek. A másik magyarázat pedig az lehet, hogy míg az első kettő tudományág jelentős része *normatív* (előíró, például gazdálkodás-, szervezés-, közgazdaságtudományok) és *deskriptív* (leíró, például állam- és jogtudományok) kutatást folytat, addig az utóbbi inkább ezekre épülő kutatást végez. Ez utóbbi esetéhez hasonlók a közigazgatástudományok, a rendészettudomány, valamint a politikatudományok is. A hadtudományok csoportja jelenthet ez alól kivételt, amely már olyan szinten specifikus tudományág, hogy saját csoportjuk tagjainak körében marad a tudományos munka szerves része.

A hálózat átmérője szerint a *politikatudományok* rendelkezik a legnagyobb értékkel (19), majd ezt követi a *közgazdaság-tudományok* és *szociológiai tudományok* köre 14-14-es értékkel. A legösszehúzóbb csoportokat a *média- és kommunikációs tudományok* (5), valamint a hadtudományok, regionális tudományok és rendészettudomány alkotják egyaránt 8-as értékkel. A *legrövidebb utak* a tagok között a *média- és kommunikációs tudományok* (2), a *hadtudományok* (3), *regionális tudományok* (3) és *rendészettudomány* (3) esetében figyelhetők meg, míg a legnagyobb értékkel kimagaslóan a *politikatudományok* tudományága rendelkezik. Ez szintén a csoportok lineáris-centrikus felépítésére utal, míg a politikatudományok inkább lineárisan szerveződő tudományág, amelynek tagjai csak egy-egy kiemelkedő szereplő által, közvetetten érintkeznek, addig a másik három sokkal szorosabb, zártabb egységet alkot.

A *közelség* értéke a legnagyobb a *média- és kommunikációs tudományok* (0,431), *a regionális tudományok* (0,300) és a *hadtudományok* (0,296), míg a legkisebb a *politikatudományok* (0,127), valamint a *közgazdaság-tudományok* (0,199) és a *szociológiai tudományok* (0,208) esetében (lásd: 4. táblázat).

4. táblázat

A közelségcentralitás (CC) értékei tudományáganként a társadalomtudomány területén

Sorszám	Tudományág	Átlagos közelség 	Név	Legmagasabb közelség 
1.	Média- és kommunikációs tudományok	0,431	Bokor Tamás	0,600
2.	Regionális tudományok	0,300	Rechnitzer János	0,483
3.	Hadtudományok	0,296	Padányi József	0,450
4.	Rendészettudomány	0,280	Varga János	0,408
5.	Közigazgatás-tudományok	0,258	Cserny Ákos	0,408
6.	Állam- és jogtudományok	0,250	Chronowski Nóra	0,333
7.	Gazdálkodás- és szervezéstudományok	0,249	Rechnitzer János	0,340
8.	Szociológiai tudományok	0,208	Murányi István	0,302
9.	Közgazdaság-tudományok	0,199	Tölgyessy Péterné Sass Magdolna	0,282
10.	Politikatudományok	0,127	Ágh Attila	0,167

Forrás: a szerzők szerkesztése

A *köztességcentralitás* alapján a *média- és kommunikációs tudományok* (0,178), valamint a *rendészettudomány* (0,052) és a *közigazgatás-tudományok* (0,050) tűnnek ki (lásd: 5. táblázat). A gazdálkodás- és szervezéstudományok esetében 0,004, az állam- és jogtudományok körében 0,009 és a szociológiai tudományok képviselői között 0,015 ez az érték.

5. táblázat

A köztességcentralitás (BC) értékei tudományáganként a társadalomtudomány területén

Sorszám	Tudományág	Átlagos köztesség 	Név	Legmagasabb köztesség 
1.	Média- és kommunikációs tudományok	0,178	Bokor Tamás	0,667
2.	Rendészettudomány	0,052	Varga János	0,316
3.	Közigazgatás-tudományok	0,050	Cserny Ákos	0,478
4.	Politikatudományok	0,040	Vitári Zsolt	0,297
5.	Hadtudományok	0,027	Padányi József	0,298
6.	Regionális tudományok	0,022	Rechnitzer János	0,484
7.	Közgazdaság-tudományok	0,020	Tölgyessy Péterné Sass Magdolna	0,230
8.	Szociológiai tudományok	0,015	Murányi István	0,234
9.	Állam- és jogtudományok	0,009	Chronowski Nóra	0,101
10.	Gazdálkodás- és szervezéstudományok	0,005	Rechnitzer János	0,086

Forrás: a szerzők szerkesztése

Az elemzés következő lépéseként érdemes a tudományterületen belüli és kívüli súlyozott kapcsolatok számát megvizsgálni. Az összes belső kapcsolat számát tekintve a legkevesebbel a *média- és kommunikációs tudományok* rendelkeznek (276 darab), majd a *közigazgatás-tudományok* (512 darab) és a *rendészettudomány* (530 darab) következik. A legtöbb belső kapcsolattal a *gazdálkodás- és szervezéstudományok* (30314 darab), a *szociológiai tudományok* (7494 darab), valamint az *állam- és jogtudományok* (6014 darab) bírnak. Ebből egyértelműen kitűnik, hogy a doktori iskolák számához és a tagok létszámához is jól illeszkedik a társszerzőségben született közlemények darabszáma.

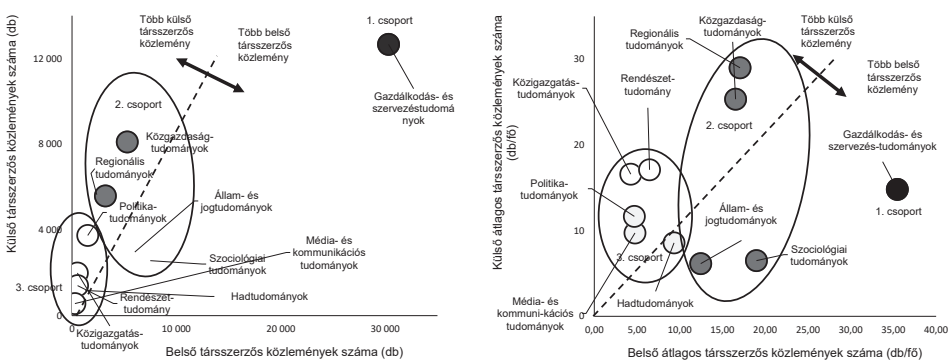
A külső kapcsolatok számában szintén a legkevesebbet mutatja a *média- és kommunikációs tudományok* csoportja (564 darab), majd a *hadtudományok* (1174 darab) és a *rendészettudomány* (1402 darab). A legtöbb külső kapcsolattal a *gazdálkodás- és szervezéstudományok* (12 662 darab), a *közgazdaságtudományok* (8107 darab) és a *regionális tudományok* (5568 darab) büszkélkedhetnek.

Ha a belső és a külső kapcsolatok összes és átlagos számát megvizsgáljuk, a vizsgált tudományágak klaszterekre bonthatók. A klaszterek számát hierarchikus klaszterezéssel határoztuk meg. Esetünkben a 10 tudományág 3 klaszterre bontható. A nem hierarchikus klaszterezésnél a K-közép-módszer használatával az alábbi három csoportot definiálhatjuk (lásd: 5. ábra):

1. Magas belső és átlagos külső kapcsolattal rendelkező csoport (1. csoport: gazdálkodás- és szervezéstudományok)

2. Átlagos belső kapcsolattal rendelkező csoport (2. csoport: regionális, közgazdaság-, állam- és jog-, illetve szociológiatudományok)

3. Alacsony belső és külső kapcsolattal rendelkező csoport (3. csoport: közigazgatás-, rendészet-, politika-, média- és kommunikációs, illetve hadtudományok)



5. ábra

A tudományágak csoportosítása a belső és külső súlyozott kapcsolatok száma alapján

Forrás: az Országos Doktori Tanács adatai és az MTMT adatbázisa alapján a szerzők szerkesztése

A belső és külső kapcsolatok arányának tekintetében vannak olyan tudományágak, amelyek inkább a belső, míg mások a külső publikálást preferálják. A legzártabb egységet a *szociológiai tudományok,* a *gazdálkodás- és szervezéstudományok,* valamint az *állam-*

és jogtudományok mutatják. Ezek esetében kétszer annyi belső kapcsolatból származó közleményt számolhatunk, mint külső együttműködésből születőt, amit azzal is indokolhatunk, hogy e tudományágak esetében a szerzők számos más doktori iskolával és azok tagjaival is kapcsolatban állnak, de a saját tudományágukon belül is találhatnak társszerzőt. A *közigazgatás-tudományok,* a *rendészettudomány,* a *politikatudományok,* valamint a *média- és kommunikációs tudományok* viszont sokkal nyitottabb tudományágak, ahol a szerzők szívesen írnak a tudományágukon kívüli társszerzőkkel, vagy azért, mert csak egyetlen doktori iskola létezik a saját területükön, tehát viszonylag szűk a mozgásterük, vagy mert tudományáguk szorosan kapcsolódik más tudományágakhoz is.

A tudományágon belüli társszerzővel írt közlemények átlagos számát tekintve szintén azt láthatjuk, hogy a legtöbb ilyen publikációval a *gazdálkodás- és szervezéstudományok* (35,45 darab/fő), a *szociológiai tudományok* (18,97 darab/fő), valamint a *regionális tudományok* (17,05 darab/fő) rendelkeznek, míg a legkevesebbet a *közigazgatás-tudományok* (4,27 darab/fő), a *politikatudományok* (4,69 darab/fő) és a *média- és kommunikációs tudományok* (4,76 darab/fő) tudhatják magukénak. A külső kapcsolatokból származó műveket tekintve viszont a *regionális tudományok* (29 darab/fő) veszik át a vezető szerepet, ezt követik a *közgazdaság-tudományok* (25,3 darab/fő), valamint a *rendészettudomány* (17,1 darab/fő), míg a legkevesebbet ilyen együttműködésben az *állam- és jogtudományok* (6,1 darab/fő), a *szociológiai tudományok* (6,5 darab/fő) és a *hadtudományok* (8,6 darab/fő) körében írnak. Ebből abszolút kitűnik, hogy a tudományágak nyitottságukat tekintve eltérnek egymástól. Míg az egyetlen doktori iskolával rendelkező, specifikus tudományágak képviselői nyitva állnak a más tudományágakhoz tartozó társszerzőkkel való együttműködés tekintetében, addig a nagyobb, több tudományos műhelyt tömörítő tudományágak több szereplőből álló, zárt közösségeket alkotnak.

A társadalomtudományi doktori iskolák jellemzése

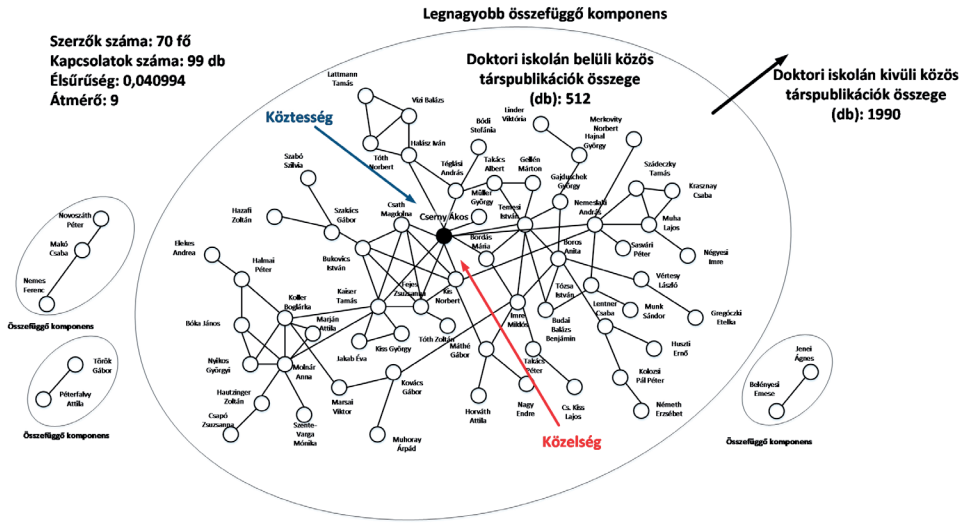
Az ODT-adatbázis szerint 2017-ben 430 darab társadalomtudományi doktori iskola működött, amelyek közül több két tudományági besorolással is rendelkezik. A kiválasztott doktori iskolák Magyarország számos városából, összesen 19 egyetemről kerülnek ki. A legtöbbet a *Budapesti Corvinus Egyetem (BCE)* működteti (6 darab), ezt követi a *Pécsi Tudományegyetem (PTE)* 5 darabbal. Említésre méltó még a három iskolát tömörítő *Eötvös Loránd Tudományegyetem (ELTE),* valamint a szintén három doktori iskolával rendelkező *Nemzeti Közszolgálati Egyetem (NKE).* A legjellemzőbb viszont az egy-egy doktori iskola működtetése társadalomtudományi területen. Az egyetemek közül 10 budapesti székhelyű, 9 vidéki intézmény.

A DI-k *csomópontjaik* számát tekintve igen különbözők. A legnépesebbek meghaladják a 100 oktatói tagot, míg a legkisebb mindössze 10 személyből áll. A legnagyobb DI a *Gazdálkodástani, BCE* 135 fővel, ezt követi *a Hadtudományi, NKE* (106 fő) (lásd: 8. ábra) és a *Gazdálkodás- és Szervezéstudományok, SZIE* (106 fő). A legkisebbek a *Vállalkozás- és Gazdálkodás, BGE* (10 fő), a *Közgazdaságtudományi, KEE* (12 fő), valamint a *Politikaelméleti, PPKE* (16 fő). A DI-k létszámának megoszlásában azt figyelhetjük meg, hogy a politikatudományi doktori iskolák inkább kisebb doktori iskolákba tömörülnek, míg a gazdálkodás- és szervezéstudományok doktori iskolái a legnépesebbek.

A csomópontok számához igazodik a *kapcsolatok* számának alakulása is. A legtöbb, 519 darab kapcsolatot a *Gazdálkodástani, BCE* doktori iskolához köthetjük, ezt követi lemaradva a *Gazdálkodás és Szervezéstudományok, SZIE* (437 darab), majd az *Ihrig Károly Gazdálkodás- és Szervezéstudományok, DE* (280 darab) következik. A legkevesebb kapcsolattal a *Vállalkozás- és Gazdálkodás, BGE* (9 darab) rendelkezik, majd a *Közgazdaságtudományi, KEE* (11 darab) és *Politikaelméleti, PPKE* (12 darab) követik. 14 DI rendelkezik 100-nál több kapcsolattal, míg 5 darab 20-nál kevesebbel. A többi pedig e két véglet között oszlik el, általában 80-100 darab közötti értékekkel. Az ezekből a kapcsolatokból a doktori iskola köreiben megszületők (belső) társközlemények számában is kiemelkedik a *Gazdálkodástani, BCE* (7856 darab), majd az *Ihrig Károly Gazdálkodás- és Szervezéstudományok, DE* (4172 darab), harmadik helyen a *Gazdálkodás- és Szervezéstudományok, SZIE* (4142 darab). A legkevesebb belső közleménnyel a *Politikaelméleti, PPKE* (32 darab), a *Politikatudományi, ELTE* (82 darab) és az *Állam- és Jogtudományi, SZE* (104 darab) rendelkezik (lásd: 6. táblázat). A külső kapcsolatokból megszülető közlemények számát tekintve is egyértelmű a *Gazdálkodástani, BCE* fölénye (13 651 darab), itt ugyanakkor a második helyen a kapcsolatok számában közepesen teljesítő *Általános és Kvantitatív Közgazdaságtan, BCE* (7674 darab) áll, harmadik helyre szorítva a *Gazdálkodás- és Szervezéstudományok, SZIE-t* (7632 darab). A legkevesebb külső közlemény a *Közgazdaságtudományi, KEE* (180 darab), az *Interdiszciplináris, ANNYE* (218 darab) és a *Politikaelméleti, PPKE* (447 darab) DI-khez kapcsolódik.

A *legnagyobb összefüggő komponensek* feltérképezése az ismeretségi hálózatokat is mutatja. Kitűnik, hogy négy doktori iskolánál mindenki egy társpublikációs hálózatban van, tehát ezeken belül nem szakadnak le kisebb csoportok, társszerzői körök. Ezek a *Mentális egészségtudományok, SE,* a *Marton Géza Állam- és Jogtudományi, DE,* az *Ihrig Károly Gazdálkodás- és Szervezéstudományok, DE,* valamint az *Állam- és Jogtudományi, PTE* doktori iskola. Közel van még a teljes csoportalkotáshoz a *Gazdálkodás- és Szervezéstudományok, SZIE,* mert 106 tagja közül 104 oktató alkot összefüggő komponenst. A legtöredezettebb képet a *Demográfia és Szociológia, PTE* mutatja, amely legnagyobb csoportjának létszáma a teljes tagsági kör 33%-át jelenti mindössze. Ezt követi a *Politikatudományi, ELTE* (37%), valamint a *Társadalmi Kommunikáció, BCE* (38%). A többi doktori iskola legnagyobb összefüggő csoportja az összes tag minimum felét tömöríti magában, de számottevő a 90%-on felüli arány (25 doktori iskola esetében).

A doktori iskolán belüli *élsűrűség* tekintetében a *Vállalkozás- és Gazdálkodás, BGE* (20%) emelkedik ki, ezt a *Közgazdaságtudományi, KEE* (16,7%), valamint az *Interdiszciplináris, ANNYE* (16,2%) követi. A legkisebb sűrűséggel pedig sorrendben a *Közigazgatás-tudományi, NKE* (4,1%) (6. ábra), a *Hadtudományi, NKE* (4,3%) és az *Állam- és Jogtudományi, ELTE* (4,6%) rendelkeznek. Egyértelműen látszik, hogy az NKE doktori iskolái között a legkisebb a sűrűség, tehát a tagok közötti potenciális kapcsolatok és a valóságban is létrejött kapcsolatok itt mutatják a legnagyobb hiányt. A *Rendészettudományi, NKE* (5,5%) az 5. legkisebb értékkel rendelkezik, míg a másik két NKE doktori iskola a két sereghajtó a sűrűség alapján.



6. ábra

A Közigazgatás-tudományi Doktori Iskola társpublikációs hálója

Forrás: az Országos Doktori Tanács adatai és az MTMT adatbázisa alapján a szerzők szerkesztése

A *klaszterezettségi együtthatóval* kapcsolatban a doktori iskolák széles skálán mozognak, 0 és 53% közötti értékekkel. A rangsort a *Vállalkozáselmélet és Gyakorlat, ME* vezeti 53%-os értékkel, ezt a *Gazdálkodás- és Szervezéstudományi, BME* (50%), valamint a *Gazdálkodástani, BCE* (49%) követi. A legkisebb értékkel egyértelműen a *Politikaelméleti, PPKE* (0%) doktori iskola rendelkezik, de a *Demográfia és Szociológia, PTE* (17%), az *Állam- és Jogtudományi, SZE* (18%) és a *Közgazdaságtudományi, KEE* (19%) szintén 20%-nál alacsonyabb értékekkel van jelen. Kitűnik a gazdálkodás- és szervezéstudományok területén működő doktori iskolák fölénye, amiből arra következtethetünk, hogy ezekben a közösségekben van a legnagyobb esélye a társszerzőségi kapcsolatok kialakulásának is. Ez szintén alátámasztja azt a tudományági összevetésnél is megállapított tendenciát, miszerint a gazdálkodás- és szervezéstudományok képviselői jól ismerik egymást, és aktív társszerzői kapcsolatokat ápolnak nemcsak a tudományáguk tagjaival, de a doktori iskoláikon belül is. A zárt közösséget alkotó, inkább belül író *Deák Ferenc Állam- és Jogtudományi, ME* szintén az élmezőnyben foglal helyet.

Az *átmérő* szerinti értékelésben az *Állam- és Jogtudományi, ELTE* (11), valamint az *Általános és Kvantitatív Közgazdaságtan, BCE* (11) kerülnek előtérbe, amely doktori iskolák egyébként nem a legnagyobb létszámúak, tehát kitűnik lineáris felépítésük. Rajtuk kívül 10-es átmérővel következnek a *Gazdálkodás- és Szervezéstudományok, KE* és a *Nemzetközi Kapcsolatok Multidiszciplináris, BCE* doktori iskolák. A legkisebb átmérőkkel a legkisebb létszámú doktori iskolák rendelkeznek, ami viszont nem meglepő eredmény. A legtöbb tagot számláló *Gazdálkodástani, BCE* (135 fő) doktori iskola mindössze 7-es átmérőjű, ami viszont jól hangsúlyozza a tagok közötti aktivitást. A minél kisebb átmérő jelentéséből eredően a lineáris kapcsolati láncolatokat hivatott megjeleníteni, amiből

következik, hogy az aktívabban kapcsolódó, a tagjai között több kapcsolatot mutató doktori iskolák inkább centrálisan épülnek fel.

Ehhez kapcsolódik a *legrövidebb utak hossza* is, amiből láthatjuk, hogy a *Politikatudományi, ELTE* közösségén belül mindenki jellemzően 1 kapcsolatra van egymástól, míg a legtöbb is 4 kapcsolatra. Ezek az *Állam- és Jogtudományi, ELTE,* az *Általános és Kvantitatív Közgazdaságtan, BCE* és a *Közigazgatás-tudományi, NKE*. A többiben inkább a 3-as legrövidebb utak a jellemzők (18 darab doktori iskola), míg a maradék 16-ban a 2 kapcsolati távolság mutatható ki.

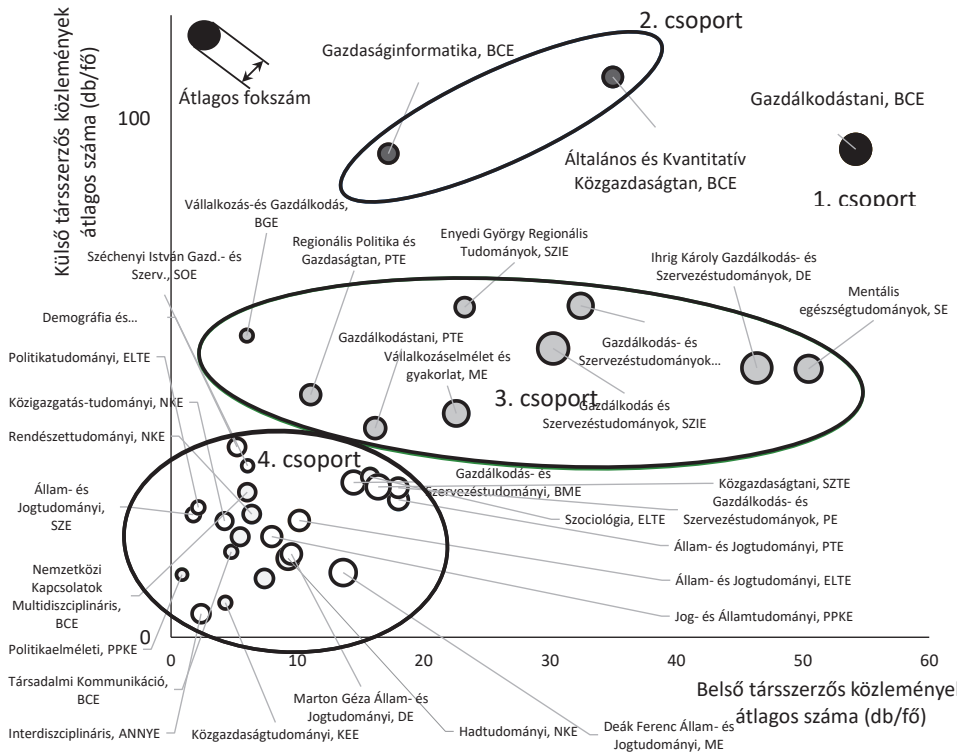
A *centralitásértékek* tekintetében a legmagasabb *közelségi értékkel* a *Politikatudományi, ELTE* (0,54), valamint a *Vállalkozás- és Gazdálkodás, BGE* (0,50) és a *Demográfia és Szociológia, PTE* (0,49) rendelkeznek. Az *Állam- és Jogtudományi, ELTE* (0,24), az *Általános és Kvantitatív Közgazdaságtan, BCE* (0,24), valamint a *Közigazgatástudományi, NKE* (0,26) pedig a leglazább összetettségű doktori iskolai hálózatok, amelyekben a tagok lazán kapcsolódnak egymáshoz. A közelségi centralitásértékekhez hasonlóan a *köztességi centralitás* értékei is nagyban befolyásolják a műhelyek hálózatának felépítését. Ebből kitűnik, hogy a *Demográfia és Szociológia, PTE* (0,28), a *Vállalkozás- és Gazdálkodás, BGE* (0,27) és a *Politikaelméleti, PPKE* (0,20) DI tagjai általában az összekötőként funkcionáló kutatók. A lista másik végén pedig a *Gazdálkodástani, BCE* (0,017), a *Gazdálkodás- és Szervezéstudományok, SZIE* (0,018), valamint az *Ihrig Károly Gazdálkodás- és Szervezéstudományok, DE* (0,022) állnak.

A 38 doktori iskola bemutatásához és összevetéséhez szükséges volt különböző statisztikai klaszterelemzések elvégzése, amelyhez kiindulópontként a hierarchikus klaszteranalízist, majd K-közép-módszert választottunk több változó alapján, amelyek közül 3 volt szignifikáns: átlagos fokszám, átlagos belső és külső társszerzős közlemények. A kapott eredmények azt mutatták, hogy négy különböző klaszter határozható meg a doktori iskolák körében (lásd: 7. ábra).

A négy kategória a következő:

- kívül és belül egyaránt számos kapcsolattal rendelkező intézmények köre (1. csoport);
- kívül sok, belül átlagos számú kapcsolattal rendelkező intézmények köre (2. csoport);
- belül és kívül egyaránt átlagos számú kapcsolattal rendelkező intézmények köre (3. csoport);
- belül sok, kívül csekély számú kapcsolattal rendelkező intézmények köre (4. csoport).

Az 1. csoportban 1 darab, a 2. csoportban 2 darab, a 3. csoportban 9 darab és a 4. csoportban 26 darab doktori iskola található.



7. ábra

A társadalomtudományi doktori iskolák csoportosítása

Forrás: az Országos Doktori Tanács adatai és az MTMT adatbázisa alapján a szerzők szerkesztése

Az 1. csoportban a *Gazdálkodástani, BCE* doktori iskolát találjuk, amely értékeivel gyakorlatilag egyedülálló a 38 másik között. A tagok átlagos kapcsolatainak száma 7,7 kapcsolat/fő, ami jelentős kutatói aktivitást és összekötöttséget mutat a tagok körében. A belső közlemények átlagos száma 54,2 darab/fő, míg kívül 94,15 darab/fő. Ez azt jelenti, hogy mind belső, mind külső körökben a kapcsolatokból jelentős publikációs teljesítmény származik, kiváló kutatói munka folyik. A doktori iskola számos más mutatóban is kiemelkedik, mint ahogy azt már a fentebbi elemzésekből is kiolvashattuk, a gazdálkodás- és szervezéstudományok tudományág amúgy is igen aktív doktori iskolái között is élen jár ilyen tekintetben.

A 2. csoport mindössze két doktori iskolát foglal magában, amelyekre a jelentős külső és átlagos belső aktivitás jellemző. A csoport két tagja a *Gazdaságinformatika, BCE* és az *Általános és Kvantitatív Közgazdaságtan, BCE.* A kapcsolatszám tekintetében az utóbbi (3,62 kapcsolat/fő) vezet, míg a másik 3,5 kapcsolat/fő átlagos értékkel bír. A belső és külső közlemények tekintetében is az *Általános és Kvantitatív Közgazdaságtan, BCE* dominál, 35 darab/fő belső közleményszámmal és 108 darab/fő külső közleménynyel. A *Gazdaságinformatika, BCE* ugyanakkor 17,2 darab/fő belső, valamint 93,3 darab/fő külső közleménnyel rendelkezik. Azt megfigyelhetjük, hogy míg a belső közlemények

tekintetében a *Gazdaságinformatika, BCE* a másik doktori iskola teljesítményének felét sem tudja produkálni, addig a külső közlemények terén már 86%-át teljesíti a másik közleményszámának. Ez az aránypár arra enged következtetni, hogy a *Gazdaságinformatika, BCE* inkább külső körökre nyitott, több külső együttműködésben született közleménnyel büszkélkedő doktori iskola, amelynek eredménye jól rímel a gazdálkodás- és szervezéstudományokra jellemző tudományági együttműködési tendenciára.

A 3. csoport tagjaira az átlagos külső és belső aktivitás jellemző. Ebben a csoportban egyértelműen a gazdálkodás- és szervezéstudományok doktori iskolái dominálnak, ami jól mutatja, hogy a zárt közösségük helyett inkább a külső kapcsolati tőkéjüket használják, és szélesebb spektrumban kutatnak oktatóik. Ha hozzátesszük, hogy a tudományági jellemzésből jól látszott, hogy tagjaik inkább a saját területük más szereplőivel működnek együtt, ebből kitűnik, hogy e tudományág kutatói körében a tudományági egység a legjellemzőbb összetartó forma, amely túlmutat a doktori iskolák körein. A legtöbb kapcsolattal ebben a csoportban a *Gazdálkodás- és Szervezéstudományok, SZIE* (8,25 kapcsolat/fő), az *Ihrig Károly Gazdálkodás- és Szervezéstudományok, DE* (7,78 kapcsolat/fő), valamint a *Mentális Egészségtudományok, SE* (6,29 kapcsolat/fő) doktori iskolái büszkélkednek. A legkevesebb kapcsolattal a *Vállalkozás- és Gazdálkodás, BGE* (1,8 kapcsolat/fő), az *Enyedi György Regionális Tudományok, SZIE* (3,53 kapcsolat/fő) és a *Regionális Politika és Gazdaságtan, PTE* (3,8 kapcsolat/fő) rendelkeznek. A legtöbb belső közleménnyel sorrendben a *Mentális Egészségtudományok, SE* (50,45 darab/fő), az *Ihrig Károly Gazdálkodás- és Szervezéstudományok, DE* (46,36 darab/fő) és a *Gazdálkodás- és Szervezéstudományok, KE* (32,4 darab/fő) vezetik a rangsort. A lista végén a *Vállalkozás- és Gazdálkodás, BGE* (6 darab/fő), a *Regionális Politika és Gazdaságtan, PTE* (11 darab/fő) és a *Gazdálkodástani, PTE* (16,2 darab/fő) helyezkednek el. A külső közlemények tekintetében a legproduktívabb doktori iskolák a *Gazdálkodás- és Szervezéstudományok, KE* (64 darab/fő), az *Enyedi György Regionális Tudományok, SZIE* (64 darab/fő) és a *Vállalkozás- és Gazdálkodás, BGE* (58 darab/fő). Ezen eredményekből láthatjuk, hogy habár a *Vállalkozás- és Gazdálkodás, BGE* és az *Enyedi György Regionális Tudományok, SZIE* tagjai nem rendelkeznek sok aktív kapcsolattal, az együttműködésekből nagy mennyiségű közlemény születik. Ez arra utalhat, hogy tudatosan választanak saját témájukhoz illő társszerzőket és partnereket, akikkel viszont kiválóan tudnak együttműködni. A külső közlemények szerinti ranglista utolsó helyezettjei ebben a csoportban a *Gazdálkodástani, PTE* (40,3 darab/fő), a *Vállalkozáselmélet és gyakorlat, ME* (43,2 darab/fő) és a *Regionális Politika és Gazdaságtan, PTE* (46,8 darab/fő) doktori iskolák.

A 4. csoport tagjai a zárt kutatói műhelyekből kerülnek ki, amelyek inkább a belső kapcsolataikra helyezik a hangsúlyt. A legtöbb kapcsolattal rendelkező doktori iskola ebben a klaszterben a *Deák Ferenc Állam- és Jogtudományi, ME* (6,08 kapcsolat/fő), majd ezt követi a *Közgazdaságtani, SZTE* (5,23 kapcsolat/fő) és a *Gazdálkodás- és Szervezéstudományi, BME* doktori iskola (4,8 kapcsolat/fő). A *Politikaelméleti, PPKE* (1,5 kapcsolat/fő), a *Demográfia és Szociológia, PTE* (1,67 kapcsolat/fő), valamint a *Közgazdaságtudományi, KEE* (1,8 kapcsolat/fő) a három sereghajtó. A belső közlemények tekintetében a listavezetők: *Állam- és Jogtudományi, PTE* (18 darab/fő), a *Gazdálkodás- és Szervezéstudományok, PE* (18 darab/fő) és *Közgazdaságtani, SZTE* (16,4 darab/fő), míg az utolsók a *Politikaelméleti, PPKE* (0,9 darab/fő), az *Állam- és Jogtudományi, SZE* (1,8 darab/fő) és a *Politikatudományi, ELTE* DI-k (2,2 darab/fő). Érdemes itt megfigyel-

nünk, hogy a legtöbb belső közleménnyel az állam- és jogtudományok doktori iskolája, a legkevesebb belső közleménnyel a két politikatudományi doktori iskola rendelkezik. A külső közlemények publikálásában a legproduktívabbak: *Széchenyi István Gazdálkodás- és Szervezéstudományok, SOE* (36,7 darab/fő), a *Regionális- és Gazdaságtudományi, SZE* (36 darab/fő) és a *Demográfia és Szociológia, PTE* (33,1 darab/fő). A legkevésbé produktívak pedig az *Interdiszciplináris, ANNYE* (4,5 darab/fő), a *Közgazdaságtudományi, KEE* (6,7 darab/fő) és az *Állam- és Jogtudományi, SZTE* DI-k (11,3 darab/fő). A csoport meghatározó mértékben az állam- és jogtudományok doktori iskoláiból áll, amelyek közül a *Deák Ferenc Állam- és Jogtudományi Doktori Iskola, ME* tűnik a legzártabb műhelynek.

Összefoglalás

A társadalomtudományi doktori iskolák elemzése érdekes és komplex hálózatelemzési problémát jelentett, amelynek eredményei különösen hasznosak mind a mesterképzésben tanuló hallgatók, mind a különböző egyetemek oktatói és kutatói számára. A doktori iskolák társszerzői hálózatának elemzése megmutatta, melyek a legintenzívebben együttműködő intézmények, kutatói csoportok. Az összevetés három szinten történt: a társadalomtudományok, a társadalomtudományi tudományágak, valamint a doktori iskolák szintjén. Az elemzéshez szükséges adatokat az MTMT- és az ODT-adatbázisokból nyertük, amelyek hivatalos és teljes körű adatokat szolgáltattak.

A tudományági összevetés eredményeként a gazdálkodás- és szervezéstudományok területének intenzív társszerzői hálózatát, valamint nyitott rendszerét láthattuk, míg az állam- és jogtudományok esetében a kutatók zártabb csoportjai tűntek ki. A doktori iskolák összevetéséből klaszterelemzés segítségével négy csoportot határoztunk meg:

- kívül és belül egyaránt számos kapcsolattal rendelkező intézmények köre (1. csoport);
- kívül sok, belül átlagos számú kapcsolattal rendelkező intézmények köre (2. csoport);
- belül és kívül egyaránt átlagos számú kapcsolattal rendelkező intézmények köre (3. csoport);
- belül sok, kívül csekély számú kapcsolattal rendelkező intézmények köre (4. csoport).

A 38 elemzésbe bevont doktori iskola közül a *Gazdálkodástani, BCE* doktori iskola emelkedik ki belső és külső aktivitásával.

Az empirikus kutatás eredményei oktatáspolitikai és tudományszervezési célokra is felhasználhatók. A tanulmány elején ismertetett hálózattudományokból átvett fogalmaknak társszerzőségi hálózatokra való alkalmazása egyedülálló a magyar szakirodalomban, viszont jól illeszkedik a nemzetközi szakirodalom kutatói hálózatokat vizsgáló közleményeihez. A kutatás jövőképeként a társszerzőségi hálózatokban szereplő oktatók kutatási témájuk szerinti összevetését tűztük ki célul, amely nemcsak a társszerzői kapcsolatok viszonyrendszerének feltérképezését, hanem tématerületek szerinti vizsgálatát is jelenti. Ezek alapján pedig az egyes doktori iskolák, valamint közvetetten a felsőoktatási intéz-

mények – az NKE Doktori Iskolái esetében lásd melléklet 2. ábra – kapcsolatát és legfőbb kutatási területeinek felderítését nyernénk.

Felhasznált irodalom

Barabási A.-L. (2016): *A hálózatok tudománya.* Budapest, Libri Kiadó.

Michalkó G. – Zsóka Á. (2016): Quo vadis, Gazdálkodástani Doktori Iskola? *Köz-gazdaság: tudományos füzetek,* 11. évf. 1. sz. 187–202.

Sasvári P. (2018): Társadalomtudományi doktori iskolák összetételének, publikációs teljesítményének empirikus elemzése és vizsgálata a nemzetközi láthatóság szempontjából. In Torgyik Judit szerk.: *Néhány társadalomtudományi kutatás és innováció.* Komárno, International Research Institute. 162–177.

Szabó G. – Bánszki T. – Ruzsányi L. (2002): A hazai doktorképzés átalakításának szükségességéről. *Magyar Tudomány,* 47. évf. 5. sz. 653–657.

Takács K. (2011): *Kapcsolatháló elemzés. Társadalmi kapcsolathálózatok elemzése*. BCE, Szociológia és Társadalompolitika Intézet. Elérhető: www.tankonyvtar.hu/hu/tartalom/tamop425/0010_2A_08_Kapcsolathalo_elemzes_szerk_Takacs_Karoly/index.html (A letöltés dátuma: 2019. 03. 12.)

Wasserman, S. – Faust, K. (1994): *Social Network Analysis. Methods and Applications.* Cambridge – New York, Cambridge University Press.

Hivatkozott dokumentumok

1994. évi XL. törvény a Magyar Tudományos Akadémiáról

387/2012. (XII. 19.) Korm. rendelet a doktori iskolákról, a doktori eljárások rendjéről és a habilitációról

Rövidítésjegyzék

Intézmények, egyetemek neve és rövidítése		
Sorszám	**Egyetem neve**	**Rövidítés**
1.	Andrássy Gyula Budapesti Német Nyelvű Egyetem	ANNYE
2.	Budapesti Corvinus Egyetem	BCE
3.	Budapesti Gazdasági Egyetem	BGE
4.	Budapesti Műszaki és Gazdaságtudományi Egyetem	BME
5.	Debreceni Egyetem	DE
6.	Eötvös Loránd Tudományegyetem	ELTE
7.	Kaposvári Egyetem	KE
8.	Közép-európai Egyetem	KEE
9.	Károli Gáspár Református Egyetem	KRE
10.	Miskolci Egyetem	ME
11.	Nemzeti Közszolgálati Egyetem	NKE

12.	Pannon Egyetem	PE
13.	Pázmány Péter Katolikus Egyetem	PPKE
14.	Pécsi Tudományegyetem	PTE
15.	Semmelweis Egyetem	SE
16.	Soproni Egyetem	SOE
17.	Széchenyi István Egyetem	SZE
18.	Szent István Egyetem	SZIE
19.	Szegedi Tudományegyetem	SZTE

Doktori iskolák neve és rövidítése		
Sorszám	**Hosszú név**	**Rövidítés**
1.	Állam- és Jogtudományi Doktori Iskola (Eötvös Loránd Tudományegyetem)	Állam- és Jogtudományi, ELTE
2.	Állam- és jogtudományi Doktori Iskola (Károli Gáspár Református Egyetem)	Állam- és jogtudományi, KRE
3.	Állam- és Jogtudományi Doktori Iskola (Pécsi Tudományegyetem)	Állam- és Jogtudományi, PTE
4.	Állam- és Jogtudományi Doktori Iskola (Széchenyi István Egyetem)	Állam- és Jogtudományi, SZE
5.	Állam- és jogtudományi Doktori Iskola (Szegedi Tudományegyetem)	Állam- és jogtudományi, SZTE
6.	Általános és Kvantitatív Közgazdaságtan Doktori Iskola (Budapesti Corvinus Egyetem)	Általános és Kvantitatív Közgazdaságtan, BCE
7.	Deák Ferenc Állam- és Jogtudományi Doktori Iskola (Miskolci Egyetem)	Deák Ferenc Állam- és Jogtudományi, ME
8.	Demográfia és Szociológia Doktori Iskola (Pécsi Tudományegyetem)	Demográfia és Szociológia, PTE
9.	Enyedi György Regionális Tudományok Doktori Iskola (Szent István Egyetem)	Enyedi György Regionális Tudományok, SZIE
10.	Gazdálkodás- és Szervezéstudományi Doktori Iskola (Budapesti Műszaki és Gazdaságtudományi Egyetem)	Gazdálkodás- és Szervezéstudományi, BME
11.	Gazdálkodás- és Szervezéstudományok Doktori Iskola (Kaposvári Egyetem)	Gazdálkodás- és Szervezéstudományok, KE
12.	Gazdálkodás- és Szervezéstudományok Doktori Iskola (Pannon Egyetem)	Gazdálkodás- és Szervezéstudományok, PE
13.	Gazdálkodás és Szervezéstudományok Doktori Iskola (Szent István Egyetem)	Gazdálkodás és Szervezéstudományok, SZIE
14.	Gazdálkodástani Doktori Iskola (Budapesti Corvinus Egyetem)	Gazdálkodástani, BCE
15.	Gazdálkodástani Doktori Iskola (Pécsi Tudományegyetem)	Gazdálkodástani, PTE
16.	Gazdaságinformatika Doktori Iskola (Budapesti Corvinus Egyetem)	Gazdaságinformatika, BCE
17.	Hadtudományi Doktori Iskola (Nemzeti Közszolgálati Egyetem)	Hadtudományi, NKE
18.	Humán Tudományok Doktori Iskola (Debreceni Egyetem)	Humán Tudományok, DE
19.	Ihrig Károly Gazdálkodás- és Szervezéstudományok Doktori Iskola (Debreceni Egyetem)	Ihrig Károly Gazdálkodás- és Szervezéstudományok, DE
20.	Interdiszciplináris Doktori Iskola (Andrássy Gyula Budapesti Német Nyelvű Egyetem)	Interdiszciplináris, ANNYE

21.	Interdiszciplináris Doktori Iskola (Pécsi Tudományegyetem)	Interdiszciplináris, PTE
22.	Jog- és Államtudományi Doktori Iskola (Pázmány Péter Katolikus Egyetem)	Jog- és Államtudományi, PPKE
23.	Közgazdaságtani Doktori Iskola (Szegedi Tudományegyetem)	Közgazdaságtani, SZTE
24.	Közgazdaságtudományi doktori iskola (Közép-európai Egyetem)	Közgazdaságtudományi, KEE
25.	Közigazgatás-tudományi Doktori Iskola (Nemzeti Közszolgálati Egyetem)	Közigazgatás-tudományi, NKE
26.	Marton Géza Állam- és Jogtudományi Doktori Iskola (Debreceni Egyetem)	Marton Géza Állam- és Jogtudományi, DE
27.	Mentális egészségtudományok Doktori Iskola (Semmelweis Egyetem)	Mentális egészségtudományok, SE
28.	Nemzetközi Kapcsolatok Multidiszciplináris Doktori Iskola (Budapesti Corvinus Egyetem)	Nemzetközi Kapcsolatok Multidiszciplináris, BCE
29.	Politikaelméleti Doktori Iskola (Pázmány Péter Katolikus Egyetem)	Politikaelméleti, PPKE
30.	Politikatudományi Doktori Iskola (Budapesti Corvinus Egyetem)	Politikatudományi, BCE
31.	Politikatudományi Doktori Iskola (Eötvös Loránd Tudományegyetem)	Politikatudományi, ELTE
32.	Regionális- és Gazdaságtudományi Doktori Iskola (Széchenyi István Egyetem)	Regionális- és Gazdaságtudományi, SZE
33.	Regionális Politika és Gazdaságtan Doktori Iskola (Széchenyi István Egyetem)	Regionális Politika és Gazdaságtan, PTE
34.	Rendészettudományi Doktori Iskola (Nemzeti Közszolgálati Egyetem)	Rendészettudományi, NKE
35.	Széchenyi István Gazdálkodás- és Szervezéstudományok Doktori Iskola (Soproni Egyetem)	Széchenyi István Gazdálkodás- és Szervezési, SOE
36.	Szociológia Doktori Iskola (Budapesti Corvinus Egyetem)	Szociológia, BCE
37.	Szociológia Doktori Iskola (Eötvös Loránd Tudományegyetem)	Szociológia, ELTE
38.	Társadalmi Kommunikáció Doktori Iskola (Budapesti Corvinus Egyetem)	Társadalmi Kommunikáció, BCE
39.	Vállalkozáselmélet és gyakorlat Doktori Iskola (Miskolci Egyetem)	Vállalkozáselmélet és gyakorlat, ME
40.	Vállalkozás- és Gazdálkodástudományi Doktori Iskola (Budapesti Gazdasági Egyetem)	Vállalkozás- és Gazdálkodás, BGE

Melléklet

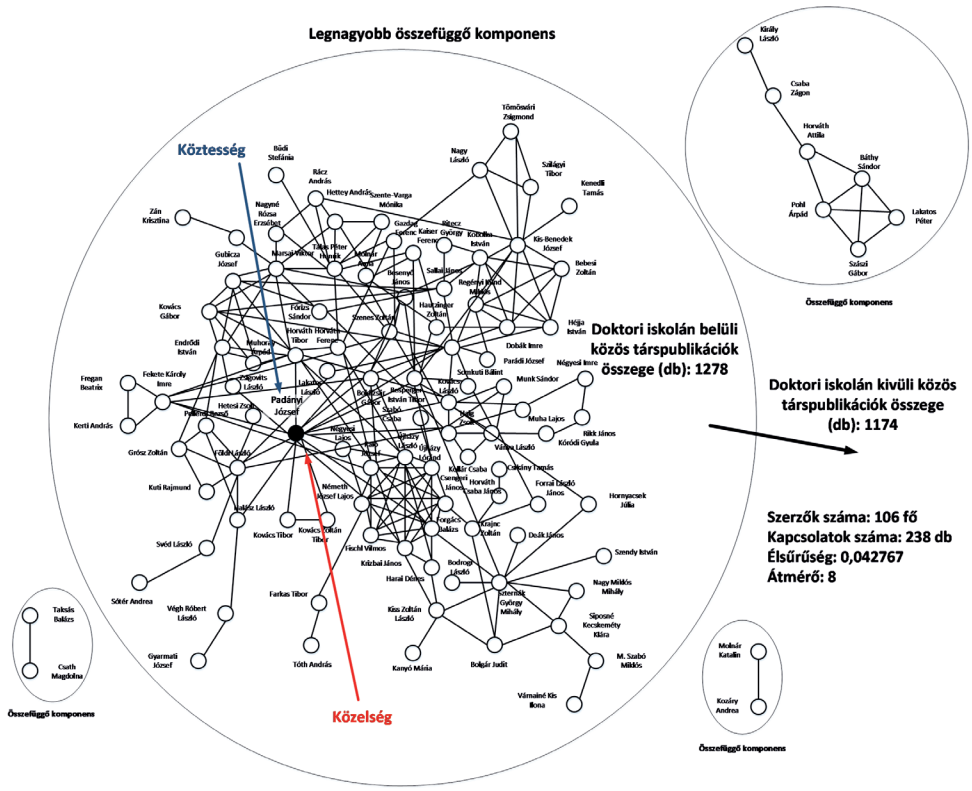
1. táblázat

A társadalomtudományi doktori iskolák tulajdonságai

Doktori Iskola neve és intézménye	Belső társközlemények száma (db)	Külső társközlemények száma (db)	Egy szerzőre jutó belső társközlemények száma (db)	Egy szerzőre jutó külső társközlemények száma (db)
Állam- és Jogtudományi, ELTE	1 076	2 388	10,15	22,53
Állam- és jogtudományi, KRE	268	950	5,47	19,39
Állam- és Jogtudományi, PTE	1 260	1 857	18,00	26,53
Állam- és Jogtudományi, SZE	104	1 396	1,76	23,66
Állam- és jogtudományi, SZTE	510	783	7,39	11,35
Általános és Kvantitatív Közgazdaságtan, BCE	2 482	7 674	34,96	108,08
Deák Ferenc Állam- és Jogtudományi, ME	940	861	13,62	12,48
Demográfia és Szociológia, PTE	218	1 191	6,06	33,08
Enyedi György Regionális Tudományok, SZIE	1 580	4 331	23,24	63,69
Gazdálkodás- és Szervezéstudományi, BME	1 112	2 299	14,44	29,86
Gazdálkodás- és Szervezéstudományok, KE	2 950	5 819	32,42	63,95
Gazdálkodás- és Szervezéstudományok, PE	1 564	2 510	17,98	28,85
Gazdálkodás és Szervezéstudományok, SZIE	4 142	7 632	30,23	55,71
Gazdálkodástani, BCE	7 856	13 651	54,18	94,14
Gazdálkodástani, PTE	1 034	2 582	16,16	40,34
Gazdaságinformatika, BCE	860	4 665	17,20	93,30
Hadtudományi, NKE	1 278	2 101	9,26	15,22
Ihrig Károly Gazdálkodás- és Szervezéstudományok, DE	4 172	4 676	46,36	51,96
Interdiszciplináris, ANNYE	114	218	2,38	4,54
Interdiszciplináris, PTE	438	1 370	4,02	12,57
Jog- és Államtudományi, PPKE	550	1 340	7,97	19,42
Közgazdaságtani, SZTE	1 098	1 944	16,39	29,01
Közgazdaságtudományi, KEE	116	180	4,30	6,67
Közigazgatás-tudományi, NKE	512	2 713	4,23	22,42
Marton Géza Állam- és Jogtudományi, DE	438	739	9,52	16,07
Mentális egészségtudományok, SE	3 330	3 418	50,45	51,79
Nemzetközi Kapcsolatok Multidiszciplináris, BCE	422	1 959	6,03	27,99
Politikaelméleti, PPKE	32	447	0,86	12,08
Politikatudományi, BCE	210	673	5,12	16,41
Politikatudományi, ELTE	82	952	2,16	25,05
Regionális- és Gazdaságtudományi, SZE	1 448	4 645	11,22	36,01
Regionális Politika és Gazdaságtan, PTE	652	2 761	11,05	46,80
Rendészettudományi, NKE	530	1 972	6,39	23,76
Széchenyi István Gazdálkodás- és Szervezési, SOE	308	2 166	5,22	36,71
Szociológia, BCE	432	1 308	7,71	23,36
Szociológia, ELTE	1 276	2 507	15,75	30,95

Doktori Iskola neve és intézménye	Belső társközlemények száma (db)	Külső társközlemények száma (db)	Egy szerzőre jutó belső társközlemények száma (db)	Egy szerzőre jutó külső társközlemények száma (db)
Társadalmi Kommunikáció, BCE	276	956	4,76	16,48
Vállalkozáselmélet és gyakorlat, ME	1 354	2 591	22,57	43,18
Vállalkozás- és Gazdálkodás, BGE	120	1 164	6,00	58,20

Forrás: a szerzők szerkesztése



1. ábra

A Hadtudományi Doktori Iskola társpublikációs hálózata

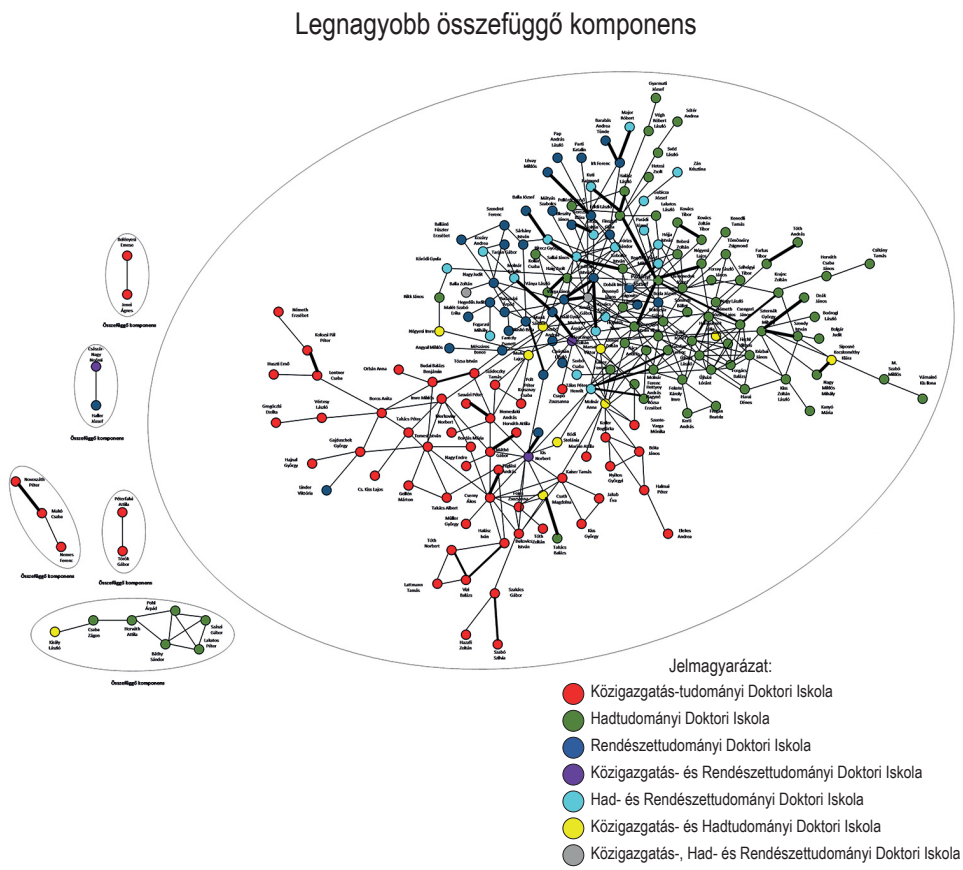
Forrás: a szerzők szerkesztése

2. táblázat

A társadalomtudományi doktori iskolák tulajdonságai

Doktori iskola neve és intézménye	**Csomópontok száma (fő)**	**Legnagyobb összefüggő komponens**	**Kapcsolatok száma (db)**	**Élsűrűség**	**Átlagos fokszám**	**Klaszterezettségi együttható**	**Átmérő**	**Legrövidebb utak hossza**	**Közöttiség**	**Közelség**
Állam- és Jogtudományi, ELTE	84	77	162	0,046	3,86	0,431	11	4	0,045	0,240
Állam- és jogtudományi, KRE	33	29	51	0,097	3,09	0,313	7	3	0,077	0,343
Állam- és Jogtudományi, PTE	48	48	90	0,080	3,75	0,306	6	3	0,045	0,336
Állam- és Jogtudományi, SZE	24	17	26	0,094	2,17	0,179	6	2	0,133	0,348
Állam- és jogtudományi, SZTE	51	40	83	0,065	3,25	0,317	7	3	0,061	0,313
Általános és Kvantitatív Közgazdaságtan, BCE	58	54	105	0,064	3,62	0,404	11	4	0,063	0,244
Deák Ferenc Állam- és Jogtudományi, ME	48	43	146	0,129	6,08	0,476	7	2	0,043	0,379
Demográfia és Szociológia, PTE	18	6	15	0,098	1,67	0,167	4	2	0,283	0,490
Enyedi György Regionális Tudományok, SZIE	47	41	83	0,077	3,53	0,284	7	3	0,060	0,307
Gazdálkodás- és Szervezéstudományi, BME	50	45	120	0,098	4,80	0,505	6	3	0,048	0,339
Gazdálkodás- és Szervezéstudományok, KE	71	67	197	0,079	5,55	0,369	10	3	0,040	0,298
Gazdálkodás- és Szervezéstudományok, PE	54	52	98	0,068	3,63	0,368	9	3	0,056	0,275
Gazdálkodás- és Szervezéstudományok, SZIE	106	104	437	0,079	8,25	0,383	8	2	0,017	0,379
Gazdálkodástani, BCE	135	129	519	0,057	7,69	0,487	7	3	0,017	0,330
Gazdálkodástani, PTE	47	45	98	0,091	4,17	0,308	7	3	0,047	0,348
Gazdaságinformatika, BCE	39	37	68	0,092	3,49	0,302	7	3	0,067	0,307
Hadtudományi, NKE	106	95	238	0,043	4,49	0,400	8	3	0,027	0,296
Ihrig Károly Gazdálkodás- és Szervezéstudományok, DE	72	72	280	0,110	7,78	0,431	6	2	0,022	0,405
Interdiszciplináris, ANNYE	21	19	34	0,162	3,24	0,398	5	2	0,082	0,433
Interdiszciplináris, PTE	59	41	95	0,056	3,22	0,337	7	3	0,054	0,337
Jog- és Államtudományi, PPKE	56	54	106	0,069	3,79	0,379	8	3	0,050	0,291
Közgazdaságtani, SZTE	47	45	123	0,114	5,23	0,349	6	2	0,038	0,396
Közgazdaságtudományi, KEE	12	8	11	0,167	1,83	0,194	4	2	0,196	0,479
Közigazgatás-tudományi, NKE	70	63	99	0,041	2,83	0,239	9	4	0,050	0,258
Marton Géza Állam- és Jogtudományi, DE	35	35	68	0,114	3,89	0,366	6	2	0,059	0,349
Mentális egészségtudományok, SE	56	56	176	0,114	6,29	0,444	6	2	0,082	0,433
Nemzetközi Kapcsolatok Multidiszciplináris, BCE	43	34	63	0,070	2,93	0,268	10	3	0,092	0,267
Politikaelméleti, PPKE	16	8	12	0,100	1,50	0,000	4	2	0,202	0,474

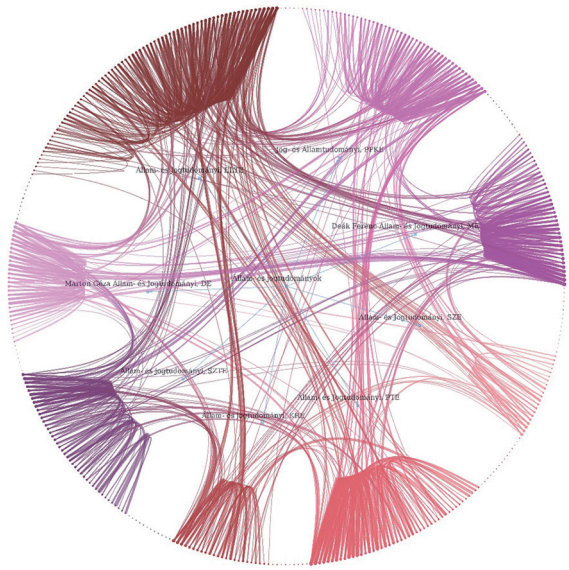
Politikatudományi, ELTE	19	7	18	0,105	1,89	0,289	4	1	0,190	0,535
Regionális- és Gazdaságtudományi, SZE	77	70	179	0,061	4,65	0,346	8	2	0,028	0,359
Regionális Politika és Gazdaságtan, PTE	44	41	84	0,089	3,82	0,292	7	3	0,057	0,321
Rendészettudományi, NKE	56	54	84	0,055	3,00	0,246	8	3	0,052	0,280
Széchenyi István Gazdálkodás- és Szervezéstudományi, SOE	34	32	51	0,091	3,00	0,234	7	3	0,075	0,324
Szociológia, BCE	36	32	55	0,087	3,06	0,367	6	2	0,061	0,365
Szociológia, ELTE	56	51	81	0,053	2,89	0,254	8	3	0,061	0,260
Társadalmi Kommunikáció, BCE	26	10	24	0,074	1,85	0,290	5	2	0,178	0,431
Vállalkozáselmélet és gyakorlat, ME	40	38	115	0,147	5,75	0,533	4	2	0,036	0,442
Vállalkozás- és Gazdálkodás, BGE	10	6	9	0,200	1,80	0,233	4	2	0,267	0,503



2. ábra

Közigazgatás-, had- és rendészettudománnyal foglalkozó kutatók társpublikációs hálózata

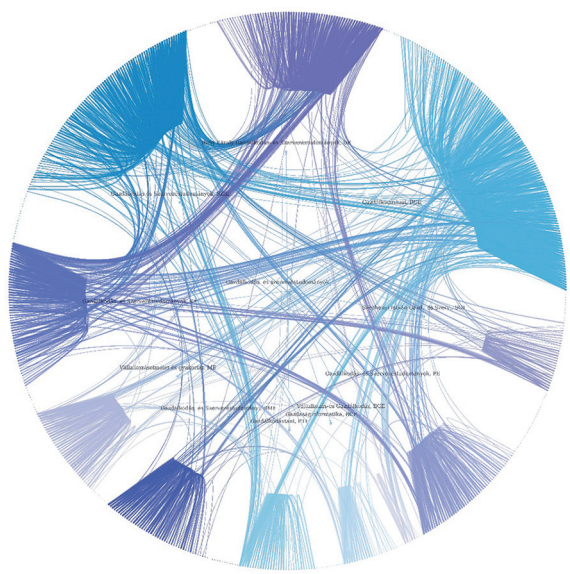
Forrás: a szerzők szerkesztése



3. ábra

Állam- és jogtudománnyal foglalkozó kutatók társpublikációs hálózata

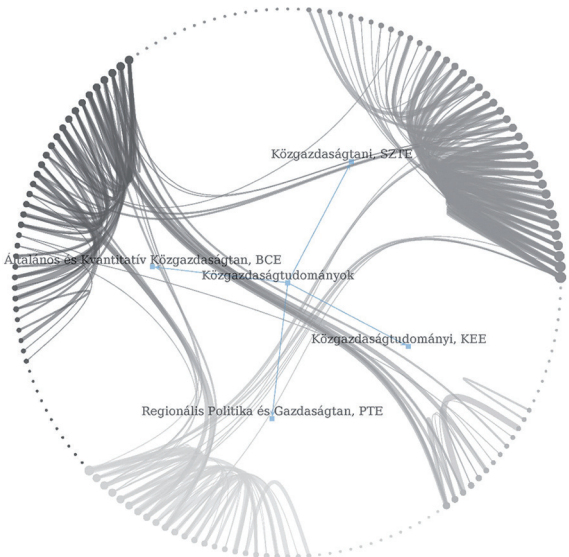
Forrás: a szerzők szerkesztése



4. ábra

Gazdálkodás- és szervezéstudománnyal foglalkozó kutatók társpublikációs hálózata

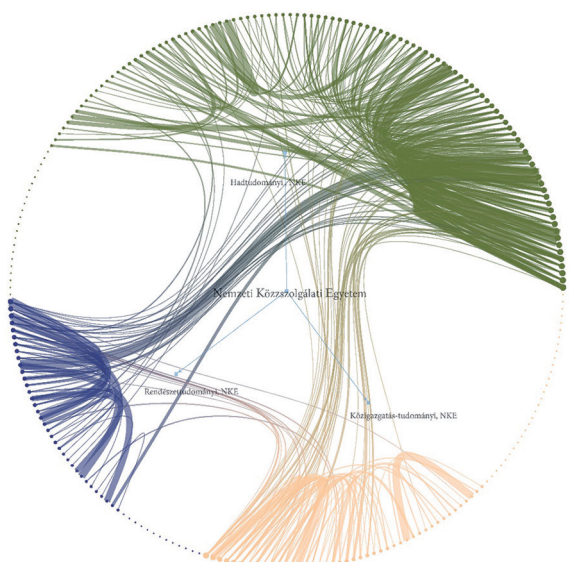
Forrás: a szerzők szerkesztése



5. ábra

Közgazdaság-tudománnyal foglalkozó kutatók társpublikációs hálózata

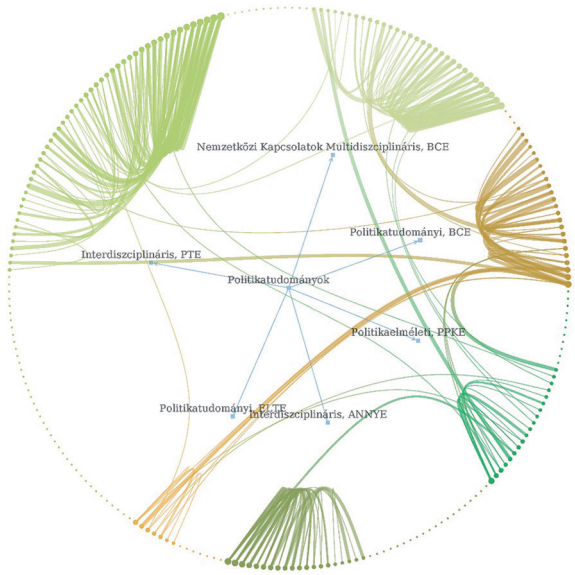
Forrás: a szerzők szerkesztése



6. ábra

Had-, rendészet- és közigazgatás-tudománnyal foglalkozó kutatók társpublikációs hálózata

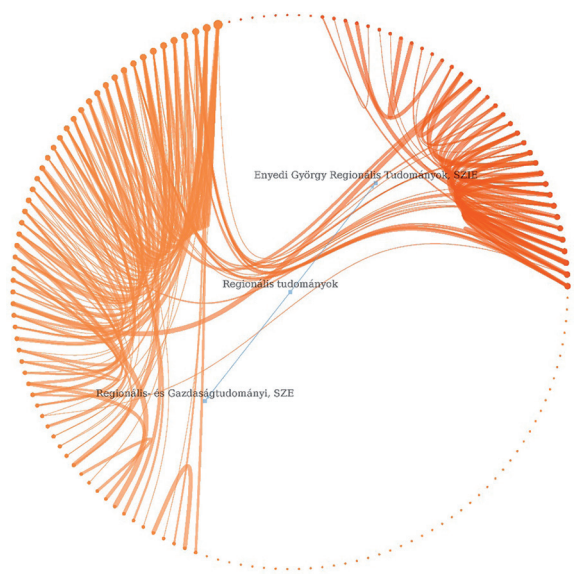
Forrás: a szerzők szerkesztése



7. ábra

Politikatudománnyal foglalkozó kutatók társpublikációs hálózata

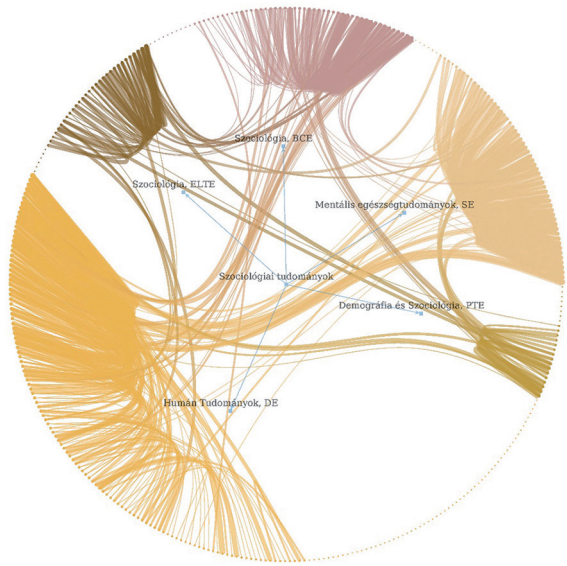
Forrás: a szerzők szerkesztése



8. ábra

Regionális tudománnyal foglalkozó kutatók társpublikációs hálózata

Forrás: a szerzők szerkesztése



9. ábra

Szociológiai tudománnyal foglalkozó kutatók társpublikációs hálózata

Forrás: a szerzők szerkesztése

Váczi Dániel – Bederna Zsolt – Szalánczi-Orbán Virág – Szádeczky Tamás

Az incidenskezelés szervezeti háttere

Bevezetés

A nemzetállamok fejlődésének mind társadalmi, mind gazdasági szempontú, pozitív irányba történő elmozdulását elősegítették a különböző információs, telekommunikációs rendszerek. Ahogy az informatika egyre meghatározóbb kezdett lenni a privát és állami szektorban működő vállalatok, szervezetek mindennapi működésében, úgy vált egyre kritikusabbá e rendszerek sértetlensége, rendelkezésre állása, az adatok bizalmassága.

Az állami kibervédelem területén a nemzeti jog és az Európai Unió NIS-irányelve is kötelezően előírja az eseménykezelő központok (Computer Emergency Response Team, a továbbiakban: CERT), illetve a számítógép-biztonsági incidenskezelő csoportok (Computer Security Incident Response Team, a továbbiakban: CSIRT) működtetését a biztonsági események kezelésére.[1] Ez a feladat tagállami és közösségi szinten egyaránt megjelenik.

A különböző nemzetállamok elkezdték kialakítani a szervezeti hátterét ennek a feladatnak, valamint folyamatosan fejlesztik a megfeleléshez szükséges képességeiket. A kibertér globalitása miatt a nemzeteknek együtt kell működni ezen a területen, ugyanakkor erről a fajta együttműködésről korlátozott mértékben érhetők el akár nemzetállami, akár uniós adatok. Mivel a hazai szervezeti együttműködésről és integrációról több releváns információ áll rendelkezésre, ezért jelen tanulmány csak a magyar helyzetet mutatja be.

Problémafelvetés

A kiberbiztonság egyik legfontosabb megoldandó problémája az információmegosztás. Különböző kezdeményezések vannak ugyan, azonban igazán működő megoldás jelenleg nincsen, hiszen az érdekelt felek – érthető módon – félnek attól, hogy megismerik az ártó szándékú támadók a rendszereiket és azok sérülékeny pontjait. Ennek ellenére már több olyan eseménykezelő központ van, amely a célnak megfelelően működik úgy, hogy számos információ elosztóközpontjaként funkcionál. Emiatt a szektor helyzete előremutatónak nevezhető. A tanulmány célja, hogy bemutassa a Magyarországon rendelkezésre álló források alapján a CERT-ek közötti kommunikáció működését. Az információáramlás legegyszerűbben hálózati ábrákon keresztül modellezhető. A hatékony kommunikáció vizsgálata során a kommunikációban részt vevő felek lesznek a csomópontok, és az élek azt fogják

[1] NIS 2016.

megmutatni, hogy közvetlenül mely szereplők tájékoztatják egymást a kiberbiztonsági incidensekre vonatkozóan.

A probléma alapjai

A 21. század technológiai újításai és a rohamos fejlődéshez idomulni próbáló társadalom új és új kihívások elé állítják az embert. A tapasztalatok azt mutatják, hogy az emberiség csupán fut a technológia, az információs ipari forradalom által előidézett komplex kihívások után, és csak nehezen tud rá reagálni. De mik is azok a problémák, amelyeket le kellene követnünk? A következőkben összegyűjtöttünk (a teljesség igénye nélkül) néhány meghatározó problémát, amelyek előidézhették az eseménykezelő központok létrejöttét.

Az információ mennyisége

Az interneten, illetve a deep és a dark weben (a továbbiakban: DDW) hatalmas mennyiségű információ áll rendelkezésre. E szempont mind a védekező, mind a támadó oldalon megjelenik a téma kapcsán. Az offenzív oldal képviselői számára (indíttatástól függetlenül) könnyű elsajátítani olyan technikákat, amelyekkel el tudják érni a céljaikat. Egyszerűen csak ismerni kell azokat a lehetőségeket, amelyek segítségével rátalálhatnak a kívánt tartalmakra, knowhow-kra. A preventív fél is számos lehetőséggel rendelkezik az információk begyűjtésére, azonban ez a hatékonyság rovására is mehet. Nekik meg kell találni azokat a hiteles, releváns forrásokat, amelyek segítségével növelni tudják védelmi képességeiket, és le tudják rövidíteni a reakcióidejüket.

Az információ változó minősége

Az előző pontban említett hatalmas információtömeg minősége természetesen nem egyenszilárdságú. Vannak hiteles források, és számos álhírrel is könnyű találkozni. (Ebben az értelemben nem az álhírek befolyásolási tényezője a fontos, hanem a megbízhatósága.) E tényező főként a védekező oldalt érinti, mivel a támadók ilyen szempontból általában lépéselőnyben vannak. Ha egy technika nem működik, addig keresnek, amíg nem találnak egy működő megoldást. Számukra az idő kisebb súllyal bír. A preventív oldal számára viszont elengedhetetlen az, hogy megfelelő minőségű, hiteles információhoz jussanak. Nem tehetik meg azt, hogy olyan forrásokból dolgozzanak, amelyek félreviszik őket, illetve hogy hamis vagy hozzá nem (teljesen) értők által készített, így torzított tartalom alapján végezzék tevékenységüket.

Az információ terjedésének gyorsasága

Ez talán az egyik legmeghatározóbb pont az összes közül, hiszen jelenlegi társadalmunk információ-központúságát is a gyors elérés mozgatja és hajtja előre. Az eseménykezelő

központok is ezen alapszanak. A különböző – ma már természetes – kommunikációs eszközök alkalmasak arra, hogy azonnal tudják értesíteni egymást, illetve a partnereiket, így növelve az érdemi reagálás hatékonyságát.

Az információhoz való hozzájutás egyszerűsége

A szemantikán alapuló, WEB 3.0-nak is nevezett információmegosztás létrejötte és annak szabad, testre szabható használata miatt a hackerek könnyen tudnak egymásnak információt átadni még úgy is, hogy arról nagy valószínűséggel reális időn belül a bűnüldöző szervek ne tudhassanak, vagy legalábbis ne tudják azonosítani – kellő pontossággal – az információ terjesztőjét. Mivel egyszerűen lehet hozzájutni információkhoz a különböző sérülékenységekről, azok kihasználhatóságának módjáról, ezért a támadók még azelőtt tudnak cselekedni, hogy a védelmi vonalak (patch-ek, IDS/IPS minták, új védelmi eszközök stb.) rendelkezésre állnának.

A határok megszűnése, a kibertérben jelen lévő nehezítő körülmények

Mi is a kibertér? Haig Zsolt és Várhegyi István *A cybertér és cyberhadviselés értelmezése* című értekezésében így fogalmaz: „Civil terminológia szerint a cybertér az elektronikus kommunikációs eszközök és rendszerek (számítógép-hálózatok, telefonvonalak, műholdas rendszerek stb.) és a rajtuk található szolgáltatások, információk alkotta virtuális tér vagy világ összefoglaló neve."[2] A kiberteret alkotó technológiák azonban nem kifejezetten egy-egy országhoz tartoznak. Határokon átívelő hálózatokról beszélünk, amelyek folyamatos kölcsönhatásban vannak egymással. A különböző anonimizáló technológiák miatt (például TOR-hálózat stb.) nem meghatározható a kommunikáló felek személyazonossága. Gyakorlatilag a kibertéren keresztül az egész világ egy egységet alkot. A védekező oldal soha sem lehet biztos abban, hogy ki a támadó fél. Ugyanakkora esély van arra, hogy a világ másik oldaláról is érkezik támadás, mint a szomszéd épületből. Emellé jön még az a tényező, hogy ugyan a fizikai síkon történő hadi vagy kriminalisztikai eredetű támadásokról sem mindig lehet 100%-osan eldönteni, hogy ki a támadó fél (például megtévesztő, befolyásoló műveletek), de a kibertámadások esetén a biztos azonosítás majdhogynem lehetetlen. Hacsak valaki magára nem vállalja, akkor igazi bizonyítékot nehéz találni. Könnyen előfordulhat az az eset is, hogy valamilyen (például hacktivista, kiberterrorista) csoport felvállal egy támadást, de ez sem jelent garanciát. Másik eshetőség, hogy egy komolyabb (például államilag támogatott) támadás esetén az eredeti tettes direkt olyan nyomokat helyez el, amelyek másra terelik a gyanút. Ilyen lehet egy bizonyos nyelv karakterkészleteinek elhelyezése a forráskódban vagy akár egy másik ország munkamenete (például időzóna, ünnepnapok figyelembevétele) alapján történő munkavégzés.

[2] Haig–Várhegyi 2008.

A felfokozott gyártói verseny, a „safety by design" hiánya

A különböző hardver- és szoftvergyártók folyamatosan versenyben vannak egymással. Arra vannak kényszerítve, hogy a piacnak megfeleljenek. Ezt úgy tudják elérni, hogy újabb és újabb termékeket hoznak ki a lehető legrövidebb idő alatt. Ez magával hordozza azt, hogy a biztonsági tesztelésre minimális idő jut. Sok esetben egyáltalán nem is merül fel tervezői oldalról, hogy az adott terméknek biztonságosnak kell lennie, így már az alapkoncepcióba sem kerül bele ez az igény. Ez az egyik oka annak, hogy számos sérülékenység marad egy-egy adott programban vagy eszközben. A folyamatos termékújítás azt is eredményezi, hogy problémás esetben már nem javítást végeznek, hanem helyette egy újabb verzió kerül a felhasználóhoz. Ezáltal a sérülékeny termékek ugyanúgy a piacon maradnak, hiszen nem hívják vissza őket maradéktalanul, csak akkor, ha valami nagyon nagy baj van. A közelmúltban ennek ellenpéldáját is láthattuk, amikor a különböző processzorokban megtalálták a Meltdown és Spectre sérülékenységet. A hibát ugyan enyhítették, azonban teljesen nem tudták a lyukat befoltozni, így maradt kitéve a legtöbb processzor e két támadási vektornak.

A különböző technikai eszközök számossága (általános és védelmi)

Fontos tényező az is, hogy az átlagos felhasználó nem tudja eldönteni a számos lehetőség közül, hogy melyik terméket érdemes megvásárolni. Egy-egy vásárlási szándék után meghozott döntés, ha egyáltalán valamennyire tudatos, általában nem teljes információn alapszik. Befolyásolja a választást az ár-érték arány, a ráfordítható anyagi keret, a korábbi tapasztalatok, az ismerősök és partnerek véleménye, de az is közrejátszhat, hogy egy adott régióban mely eszközök, szoftverek az ismertek, az elterjedtek. Nehéz kiválasztani ezek alapján azt a legjobbat, ami valószínűleg biztonságos is. De ez a kérdéskör a specifikus biztonsági termékekre is igaz. Az újabb és újabb kiberbiztonsági fenyegetettségek lefedése nagyon költséges, ráadásul a különböző területeken újabb és újabb gyártók termékei jelennek meg. Egyáltalán nem mindegy azonban e megoldások használatának, konfigurálásának a módja sem. Érdemes ezért kellő erőforrást szánni a kiválasztásra, hogy valóban a kívánt hatást érjük el a bevezetéssel.

Rabló-pandúr harc (ha a bűnözők összefognak, akkor a védelmi oldalnak sem ártana)

Az élet minden területén, ahol értelmezhető támadó és az azt megakadályozni kívánó védekező tevékenység, folyamatos harc zajlik az előnyök megszerzéséért. Nincs ez másként a kiberbiztonság területén sem. Ahhoz, hogy ebben a „versenyben" a védelem felé billenjen a mérleg, egy nagyon fontos dolgot kell kialakítani. Ez pedig az információmegosztás. Míg a támadási oldalon sok esetben ez a folyamat működik, addig a védelmi oldal még csak most kezdi a saját metódusait kialakítani. Korábban mindenki védte a saját rendszereit, azonban be kellett látnia a szakmának, hogy ez ebben a formában nem kivitelezhető, már csak ha a szakemberek hiányát vesszük is alapul. E probléma feloldására jöttek létre a kü-

lönböző eseménykezelő központok, amelyek ezt a csomóponti szerepet kívánják betölteni a számos szereplőből álló közösségben.

A nemzetközi és hazai incidensek statisztikai ismertetése

Az eseménykezelő központok és az ahhoz kapcsolódó szervezetek, szabályozások, adatszolgáltatások lehetőséget biztosítanak számunkra, hogy nemzetközi és globális elemzésekhez hozzáférjünk. Az adatok nagy mennyiségben állnak rendelkezésünkre, és különböző szervezetek időközönként jelentéseket, iránymutatásokat és biztonsági előírásokat is közzétesznek a hatékonyabb védelem elérése érdekében. Nemzetközi viszonylatban két frissebb jelentés anyagát emelnénk ki.

Microsoft-fenyegetéselemzés

A Microsoft globális fenyegetéselemzése[3] rámutat arra, hogy globálisan

- átlagosan 146 nap telik el, míg észlelik a támadókat a hálózaton,
- a kiberbűnözés összes lehetséges költsége a globális közösség számára 500 billió dollár,
- a kompromittált hitelesítő adatok teszik ki az összes hálózati befogadás 63%-át,
- az adatszegések átlagos vállalati költsége 3,8 millió dollár,
- minden ötödik kis- és középvállalkozásnak szüksége lenne kiberbűnözés elleni védelemre.

A 2018-as „Data breach" vizsgálati jelentés

A Verizon jelentésében[4] felhasznált adatok 65 országtól származnak, és 53 ezer valós, világméretű incidensből készültek, ebben az évben több mint 53 ezer, köztük 2,216 megerősített adatszegésből. A jelentésekből következtetésként levonható, hogy

- a jogsértések 76%-a pénzügyileg motivált. A legtöbb számítógépes bűnözőt motiválja, hogy valamilyen módon pénzt szerezzen tevékenysége során, ez lehet a fizetési kártyaadatok ellopása, személyesen azonosítható információ vagy szellemi tulajdon ellopása stb.,
- az internetes támadásoknak közel háromnegyede (73%) kívülről történt,
- a belső fenyegetések aránya 10% feletti,
- az emberek 4%-a bármely adathalász kampányra rákattint.

[3] Microsoft 2018.
[4] Verizon 2018.

1. táblázat
Jogsértések ágazatonként

Ágazat	Adatszegések száma	Incidensek száma	Ki?	Milyen?	Hogyan?
Szállás	338	368	99% külső, 1% belső	93%-os fizetés (POS), 5% személyes, 2% hitelesítő	93% hackelés, 91% rosszindulatú program
Oktatás	101	292	81% külső, 19% belső	72% személyes, 14% titkok, 11% orvosi	46% hackelés, 41% társadalmi
Pénzügy	146	598	79% külső, 19% belső	36% személyes, 34%-os fizetés, 13% bank	34% hackelés, 34% fizikai
Egészségügy	536	750	43% külső, 56% belső	79% orvosi, 37% személyes, 4% fizetés	35% hiba, 24% visszaélés
Információ	109	1 040	74% külső, 23% belső	56% személyes, 41% hitelesítő, 9% belső	57% hackelés, 26% hiba
Gyártás	71	536	89% külső, 13% belső	32% személyes, 30% titok, 24% hitelesítő	66% hackelés, 34% rosszindulatú program
Szakmai	132	540	70% külső, 31% belső	56% személyes, 28% hitelesítő, 16% belső	50% hackelés, 21% társadalmi
Közösségi	304	22 788	67% külső, 34% belső	41% személyes, 24% titkok 14% orvosi	52% hackelés, 32% társadalmi
Kiskereskedelem	169	317	91% külső, 10% belső	73%-os fizetés, 16% személyes, 8% hitelesítő	46% hackelés, 40% fizikai

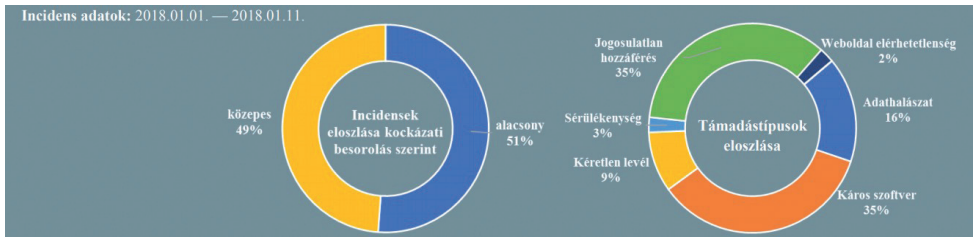
Forrás: Verizon 2018

A jogsértések 68%-át hónapokig vagy tovább tartott felfedezni, míg a támadások percekig vagy csak másodpercekig tartottak.

Jól látható, hogy az adatok és az információk ellátottsága adott, a nemzetközi incidensek típusaiból és irányultságából következtetések és fejlesztési tervek szűrhetők le. Az incidensek széles körűek, és a gazdaság, valamint a társadalom egészét érintik, így ezek kezelése komplex feladat.

Magyarországi események statisztikai elemzése

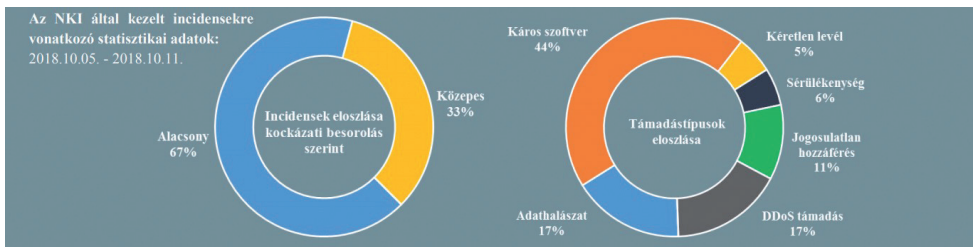
A Nemzeti Elektronikus Információbiztonsági Hatóság (NEIH) rendelkezésre bocsátja a Nemzetközi IT-biztonsági sajtószemlét, amely heti lebontásban tartalmazza az incidensekre vonatkozó legfrissebb adatokat, így lehetőség van az incidenseket kockázati besorolásuk szerinti elosztásban és támadástípusok szerinti elosztásban is elemezni.



1. ábra

Incidensek és támadások eloszlása (2018. január)

Forrás: NEIH 2018a



2. ábra

Incidensek és támadások eloszlása (2018. október)

Forrás: NEIH 2018b

Az események és az incidensek által generált és elemzett adatokból is jól látható, hogy minden szervezet és független cég vagy szakosodott intézet a saját adataiból saját értéket hoz létre, saját elemzési kritériumokkal és célokkal. Az incidensadatok alapján megállapíthatók a főbb események és ezek előfordulási aránya, változása, amely iránymutatást adhat a fejlesztendő területek és szükséges lépések megtételére. Az információmegosztás és az információk megfelelő helyre való eljuttatása a szervezetek között jelentős hatékonyságnövelést érne el az információbiztonság területén.

A fogalmak tisztázása

A téma alapját a CERT-ek és a CSIRT-ek biztosítják. A következő fejezetben bemutatjuk a kettő közti különbséget az Európai Hálózat- és Információbiztonsági Ügynökség (European Union Agency for Network and Information Security, a továbbiakban: ENISA) a csoportok létrehozásával foglalkozó kiadványa alapján.[5]

A CERT, azaz Computer Emergency Response Team (magyarul számítógépesvészhelyzet-kezelő csoport) egy olyan szervezet, amely a kiberbiztonság területén megfelelő időben és minőségben különböző szolgáltatásokat nyújt. A kifejezést az Amerikai Egyesült Államokban védette le a CERT Coordination Center (CERT/CC). Ennek európai megfelelője

[5] ENISA 2016.

a CSIRT, azaz Computer Security Incident Response Team (magyarul számítógép-biztonsági incidenskezelő csoport). Mindkét típusú szervezetnél hasonló jellegű munkát végeznek, azonban nincs egyiknél sem letisztázva, mivel kell pontosan foglalkoznia egy-egy reagáló csapatnak, iránymutatások vannak. A gyakorlatban előfordul olyan értelmezés is, miszerint a CERT egy magasabb fejletségi, érettségi szintnek felel meg, míg a CSIRT kevesebb szolgáltatást nyújt. Hálózatelemzési szempontból azonban nincs jelentősége a megkülönböztetésnek, így a továbbiakban szinonimaként kerül használatra a két szervezeti típus.[6]

Előfordul a két fő megjelölésen kívül, hogy hasonló funkciójú csoportokat más rövidítéssel is ellátnak. Ilyen lehet például az IRT, azaz Incident Response Team (magyarul incidenskezelő csoport), a CIRT, azaz Computer Incident Response Team (magyarul számítógép incidenskezelő csoport) és a SERT, azaz Security Emergency Response Team (magyarul biztonsági vészhelyzetkezelő csoport). Mindegyik csoportnak ugyanaz a feladata. A benne lévő szakemberek elsődleges feladata, hogy az ügyfeleiket ért számítógépes incidensek esetén beavatkozzanak, illetve minimalizálják az esetleges károkat.

A csoportok, attól függően, hogy milyen ügyfélkörnek, szektornak nyújtják a szolgáltatásaikat, különböző szintű és minőségű feladatokat láthatnak el. Ezek a szektorok a következők:

- tudományos szektor,
- üzleti szektor,
- kritikus (információs) infrastruktúrák szektora,
- kormányzati szektor,
- katonai szektor,
- kis- és középvállalkozások szektora.

Ezenkívül még szükséges megemlíteni, hogy lehet külön nemzeti CSIRT, illetve sokszor a szoftverkiadók vagy más cégek belső csoportokat üzemeltetnek erre a célra.

Jogszabályi ismertető

Az eseménykezelő központok létrejöttét és jelenlegi formában történő működését az igény mellett a törvényalkotók is elősegítették. Nemcsak az Amerikai Egyesült Államokban, hanem az Európai Unión belül is hamar rájöttek, hogy a kibertér fenyegetései elleni védekezés egyik sarokpontja, hogy magas szinten le tudják fektetni a védelem minél magasabb fokú eléréséhez szükséges környezetet. Magyarország az unión belül az elsők között szerepelt a kibervédelemmel kapcsolatos szabályzók megalkotásában. A CERT-ek működésével kapcsolatban a következő jogszabályok hatályosak jelenleg (2018. október):

- 65/2013. (III. 8.) Korm. rendelet a létfontosságú rendszerek és létesítmények azonosításáról, kijelöléséről és védelméről szóló 2012. évi CLXVI. törvény végrehajtásáról
- 1139/2013. (III. 21.) Korm. határozat Magyarország Nemzeti Kiberbiztonsági Stratégiájáról
- 2013. évi L. (IV. 15.) törvény az állami és önkormányzati szervek elektronikus információbiztonságáról

[6] Tikos 2018.

- 233/2013. (VI. 30.) Korm. rendelet az elektronikus információs rendszerek kormányzati eseménykezelő központjának, ágazati eseménykezelő központjainak, valamint a létfontosságú rendszerek és létesítmények eseménykezelő központja feladat- és hatásköréről
- 187/2015. (VII. 13.) Korm. rendelet az elektronikus információs rendszerek biztonsági felügyeletét ellátó hatóságok, valamint az információbiztonsági felügyelő feladat- és hatásköréről, továbbá a zárt célú elektronikus információs rendszerek meghatározásáról
- A katasztrófavédelemről és a hozzá kapcsolódó egyes törvények módosításáról szóló 2011. évi CXXVIII. törvény
- Az elektronikus ügyintézés és a bizalmi szolgáltatások általános szabályairól szóló 2015. évi CCXXII. törvény
- A nemzeti adatvagyon körébe tartozó állami nyilvántartások védelméről szóló 2010. évi CLVII. törvény
- 38/2011. (III. 22.) Korm. rendelet a nemzeti adatvagyon körébe tartozó állami nyilvántartások adatfeldolgozásának biztosításáról
- 2012. évi CLXVI. törvény a létfontosságú rendszerek és létesítmények azonosításáról, kijelöléséről és védelméről
- 484/2013. (XII. 17.) Korm. rendelet a Nemzeti Kiberbiztonsági Koordinációs Tanács, valamint a Kiberbiztonsági Fórum és a kiberbiztonsági ágazati munkacsoportok létrehozásával, működtetésével kapcsolatos szabályokról, feladat- és hatáskörükről
- 185/2015. (VII. 13.) Korm. rendelet a kormányzati eseménykezelő központ és az eseménykezelő központok feladat- és hatásköréről, valamint a biztonsági események kezelésének, a biztonsági események műszaki vizsgálatának és a sérülékenységvizsgálat lefolytatásának szabályairól

A területet magasabb, uniós szinten a következők szabályozzák:

- Az Európai Parlament és a Tanács 460/2004/EK rendelete (2004. március 10.) az Európai Hálózat és Információbiztonsági Ügynökség létrehozásáról
- 2005. Zöld Könyv a kritikus infrastruktúra védelemről
- 2009. Kritikus Információs Infrastruktúra Védelme [COM(2009) 149]
- 2010. Digitális Menetrend (Európa 2020 Stratégia része)
- 2010. Közlemény az Európai Hálózat és Információbiztonsági Ügynökség (ENISA) megerősítésére és modernizálására vonatkozóan.
- 2011. Közlemény a kritikus informatikai infrastruktúrák védelméről: „Eredmények és következő lépések: a globális kiberbiztonság felé” [COM(2011) 163]
- Az Európai Unió Kiberbiztonsági Stratégiája: egy nyílt, biztonságos és megbízható kibertér (2013)
- Az Európai Parlament és a Tanács 526/2013/EU rendelete (2013. május 21.) az Európai Uniós Hálózat- és Információbiztonsági Ügynökségről (ENISA) és a 460/2004/EK rendelet hatályon kívül helyezéséről
- Az Európai Parlament és a Tanács (EU) 2016/1148 irányelve (2016. július 6.) a hálózati és információs rendszerek biztonságának az egész Unióban egységesen magas szintjét biztosító intézkedésekről

CERT-feladatok

A tagállamok minimumképességeinek kialakítása mind az Európai Unió kiberbiztonsági stratégiájában, mind a NIS-irányelvekben megfogalmazódik. Ez utóbbinak a 9. cikke határoz arról, hogy minden tagállamnak legalább egy CSIRT-et ki kell jelölni, amely lefedi az előírt ágazatokat, valamint az előírt szolgáltatásokat.

2. táblázat

A NIS-irányelv által lefedett ágazatok

Állami és önkormányzati szervek	
Digitális szolgáltatók	
Közlekedés	közúti
	vízi
	vasúti
	légi
Energia	földgáz
	kőolaj
	villamos energia
Digitális infrastruktúra	TLD
	DNS
	IXP
Ivóvíz	ellátás
	elosztás
Egészségügyi ellátó létesítmények	
Pénzügy	pénzügyi piaci infrastruktúrák
	banki szolgáltatások

Forrás: a szerzők szerkesztése

3. táblázat

A CSIRT-szolgáltatások listája a CERT/CC-től

Válaszintézkedésként nyújtott szolgáltatások	riasztások és figyelmeztetések
	incidenskezelés
	incidenselemzés
	incidenssel kapcsolatos helyszíni válaszintézkedések
	incidenssel kapcsolatos válaszintézkedések támogatása
	incidenssel kapcsolatos válaszintézkedések koordinálása
	sebezhetőség kezelése
	sebezhetőség elemzése
	sebezhetőséggel kapcsolatos válaszintézkedések
	sebezhetőséggel kapcsolatos válaszintézkedések koordinálása

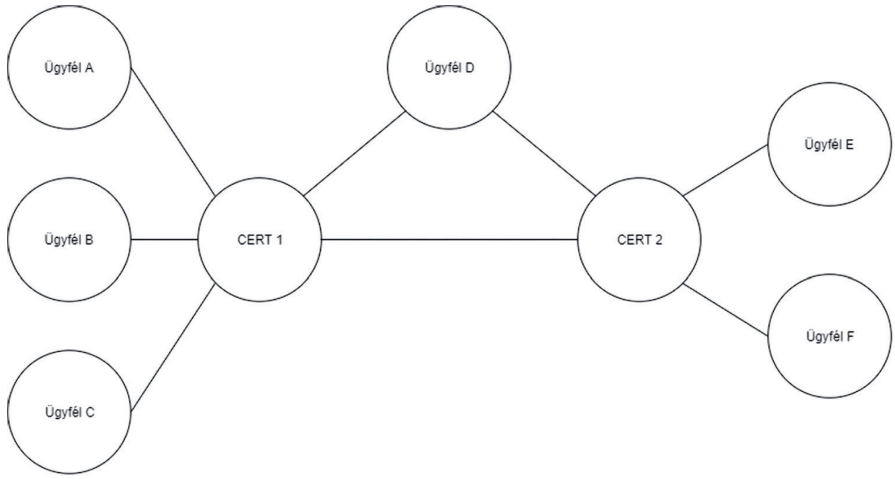
Megelőző szolgáltatások	bejelentések
	technológiafigyelés
	biztonsági ellenőrzések és felmérések
	a biztonsági konfiguráció beállítása és karbantartása
	biztonsági eszközök fejlesztése
	behatolásérzékelési szolgáltatások
	biztonsággal kapcsolatos információk terjesztése
Torzulások kezelése	torzulások elemzése
	torzulásokkal kapcsolatos intézkedések
	torzulásokkal kapcsolatos intézkedések koordinálása
A biztonsági minőségirányítás	kockázatelemzés
	üzletmenet-folytonosság és katasztrófa utáni visszaállítás
	biztonsági tanácsadás
	a tudatosság növelése
	oktatás/képzés
	termék kiértékelése vagy tanúsítása

Forrás: ENISA 2016

Az ügyfeleik incidenskezelése (a riasztás, az elemzés, a válaszlépések támogatása) egy meghatározó része a CSIRT-feladatoknak. Ezenkívül előfordulhat, hogy egy-egy csoport foglalkozik sebezhetőségmenedzsmenttel, biztonsági képességek fejlesztésével, karbantartásával, illetve a releváns információk megosztásával. Ezek közül hálózatvizsgálat szempontjából az utolsót szükséges kiemelni, amelyhez kapcsolódik a NIS-direktíva által előírt részvételi kötelezettség a CSIRT-ek hálózatában. Az információbiztonsági szakmában alapvetően hiányzik, hogy a hatékony működés érdekében minél több tudást, információt, tapasztalatot adjanak át a különböző szervezetek egymásnak. Az ilyen jellegű együttműködést érthető módon az információk érzékenysége nehezíti meg. Egyik cég sem árulja el szívesen a gyenge pontjait, mint ahogy azt sem, hogy milyen biztonsági rendszereket használ adatai, infrastruktúrájának védelmére. E probléma feloldása az egyik fő célja a CERT-eknek, CSIRT-eknek. Szükség van a szakmában olyan független szereplőkre, akik a lehető legszéleskörűbben, szakszerűen tudnak ilyen jellegű információmegosztással foglalkozni.

A magyarországi eseménykezelő központok által felrajzolható hálózat

Ez viszont akkor tud megfelelően működni, ha a CERT-ekhez beérkező információk nem vesznek el, hanem valóban eljutnak a szükséges helyekre.



3. ábra

Általános kapcsolati diagram a CERT-ek és ügyfeleik között

Forrás: a szerzők szerkesztése

Magyarországon talán a legismertebb az Ibtv. által előírt Kormányzati Eseménykezelő Központ, azaz a GovCERT. Ezen kívül azonban más hasonló funkciót betöltő intézmények is tevékenykednek. Az Országos Katasztrófavédelmi Főigazgatóság felügyelete alatt működik a Létfontosságú Rendszerek és Létesítmények Informatikai Biztonsági Eseménykezelő Központ (a továbbiakban: LRLIBEK), a katonai eseménykezelésért felelős MILCERT. A Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutatóintézetében (MTA SZTAKI) működik az Internet Szolgáltatók Tanácsának (ISZT) támogatásával létrejött Hun-CERT, valamint az elsősorban felsőoktatási és kutatási területen tevékenykedő Nemzeti Információs Infrastruktúra Fejlesztési (NIIF) Intézet incidenskezelési csoportja, a NIIF-CSIRT. A felsorolt lista természetesen nem teljes.

Ha a hálózatot kívánjuk feltérképezni, fontos ismerni, hogy kik az adott központ „ügyfelei". Mivel ezek bizalmas információk, az egyetlen kiindulási alapot az Ibtv. által a GovCERT-hez hozzárendelt szervek jelentik, amelyek a következők:

- a központi államigazgatási szervek a Kormány és a kormánybizottságok kivételével,
- a Köztársasági Elnöki Hivatal,
- az Országgyűlés Hivatala,
- az Alkotmánybíróság Hivatala,
- az Országos Bírósági Hivatal és a bíróságok,
- az ügyészségek,
- az Alapvető Jogok Biztosának Hivatala,
- az Állami Számvevőszék,
- a Magyar Nemzeti Bank,
- a fővárosi és megyei kormányhivatalok,
- a helyi és a nemzetiségi önkormányzatok képviselő-testületének hivatalai, a hatósági igazgatási társulások,
- a Magyar Honvédség.

A különböző CERT-ek és CSIRT-ek önálló működése mellett fontos megemlíteni az úgynevezett CERT/CSIRT közösségeket. Az első, az 1990-ben FIRST (Forum of Incident Response and Security Teams) néven létrejött ilyen csoportosulás óta számos, általában jogi kötelezettség nélküli, önkéntes nemzeti, regionális, szektorális alapon létrehozott együttműködés született. Tagjaik információkat osztanak meg egymással, illetve incidensek esetén együttműködnek különböző szinten, hatásfokkal, bizonyos esetekben akár hatósági, jogalkotói szereplők bevonásával együtt. Hálózatvizsgálat szempontjából az adott ország nemzeti entitásai közötti kommunikáció útját érdemes kibővíteni egy sokkal szélesebb körű háló felrajzolásával, hiszen csak a FIRST 350 feletti taggal rendelkezik. Ezenkívül pedig hazánk szempontjából érdemes szót ejteni a Forum of Incident Response Teamsről és az Internet Watch and Warning Networkről (IWWN), amelynek tagjai 15 ország központjaiból tevődnek össze, illetve a Trans-European Research and Education Networking Associationről (TERENA). Az európai közösségek közül érdemes még megemlíteni a Trusted Introducert (TI) és a 12 uniós ország kormányzati CSIRT-jéből és az CERT-EU-ból álló European Government CERTs Groupot. Magyarország tagja ezeken kívül regionális közösségeknek is, mint például a Közép-európai Kiberbiztonsági Platformnak (Central European Cyber Security Platform, CECSP).

A teljes hálózat feltérképezése érdekében szükséges említést tenni az úgynevezett az információmegosztó és -elemző központokról (Information Sharing and Analysis Center, a továbbiakban: ISAC) és az információmegosztó és -elemző szervezetekről (Information Sharing and Analysis Organisation) is, amelyek hidat képeznek a kormányzati és a privát szektor között. Jelenleg két ilyen működő szervezetről beszélhetünk európai szinten, amelyek az energetikai és a pénzügyi szektorban tevékenykednek. Ezek az European Energy – Information Sharing Analysis Centre (EE-ISAC), valamint az European Financial Institutes – Information Sharing and Analysis Centre (FI-ISAC).

CERT-ek hálózata

Az információáramlás megértéséhez érdemes a különböző szereplőket egy hálózati ábrán felrajzolni. A következő ábrán szerepel Magyarország viszonylatában öt nagyobb CERT, illetve a speciálisan hozzájuk rendelt területek. Az Ibtv. által a GovCERT-hez rendelt feladatokat már az előző alfejezetben tárgyaltuk, azonban fontos kiemelni az Országos Katasztrófavédelmi Főigazgatóság felügyelete alatt működő Létfontosságú Rendszerek és Létesítmények Informatikai Biztonsági Eseménykezelő Központhoz tartozó szervezeteket. Idesorolhatók mindazon szervezetek, amelyek a létfontosságú rendszerek és létesítmények azonosításáról, kijelöléséről és védelméről szóló 2012. évi CLXVI. törvény mellékleteiben meghatározott ágazatokhoz és alágazatokhoz tartoznak. Ezek a következők:

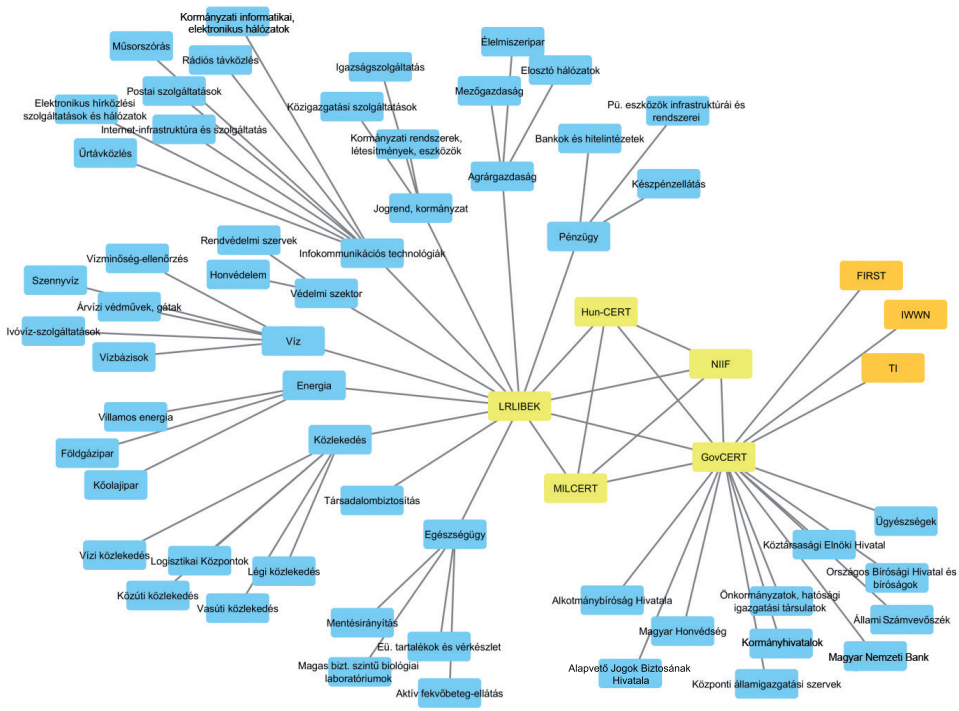
4. táblázat

Létfontosságú ágazatok és alágazatok

Ágazat	**Alágazat**
Energia	a villamosenergia-rendszer létesítményei (kivéve az atomerőmű nukleáris biztonságára és sugárvédelmére, fizikai védelmére, valamint biztosítéki felügyeletére vonatkozó szabályozás hatálya alá tartozó rendszerek és rendszerelemek)
	kőolajipar
	földgázipar
Közlekedés	közúti közlekedés
	vasúti közlekedés
	légi közlekedés
	vízi közlekedés
	logisztikai központok
Agrárgazdaság	mezőgazdaság
	élelmiszeripar
	elosztóhálózatok
Egészségügy	aktív fekvőbeteg-ellátás
	mentésirányítás
	egészségügyi tartalékok és vérkészletek
	magas biztonsági szintű biológiai laboratóriumok
	gyógyszer-nagykereskedelem
Társadalombiztosítás	társadalombiztosítási ellátások igénybevételéhez kapcsolódó informatikai rendszerek és nyilvántartások
Pénzügy	pénzügyi eszközök kereskedelmi, fizetési, valamint klíring- és elszámolási infrastruktúrái és rendszerei
	bank- és hitelintézeti biztonság
	készpénzellátás
Infokommunikációs technológiák	internet-infrastruktúra és internethozzáférés-szolgáltatás
	vezetékes és vezeték nélküli elektronikus hírközlési szolgáltatások, vezetékes és vezeték nélküli hírközlő hálózatok
	rádiós távközlés
	űrtávközlés
	műsorszórás
	postai szolgáltatások
	kormányzati informatikai, elektronikus hálózatok
Víz	ivóvíz-szolgáltatás
	felszíni és felszín alatti vizek minőségének ellenőrzése
	szennyvízelvezetés és -tisztítás
	vízbázisok védelme
	árvízi védművek, gátak
Jogrend – Kormányzat	kormányzati rendszerek, létesítmények, eszközök
	közigazgatási szolgáltatások
	igazságszolgáltatás
Közbiztonság – Védelem	rendvédelmi szervek infrastruktúrái
Honvédelem	honvédelmi rendszerek és létesítmények

Forrás: a szerzők szerkesztése

A teljes hálózat feltérképezéséhez szükség lenne azoknak a szervezeteknek a listájára, amelyeket valóban besoroltak a létfontosságú rendszerelemek közé. Ez az információ azonban nem nyilvános, így a hálózati ábrán ágazat-alágazat bontásban kerültek megjelenítésre. További kibontásra lenne szükség a GovCERT alá tartozó szervezetekhez, azonban a kormányzati berendezkedés folyamatos változása miatt ez sem történt meg. Ettől függetlenül egy fejlesztési projekt keretén belül, ahol az optimalizálás a cél, fontos lehet az aktuális helyzet feltérképezése. Az ábrában tehát a kékkel jelölt entitások egy közelítő, sok esetben összefoglaló csoportot jelölnek. Citromsárgával tüntettük fel az öt talán legnagyobb központi jelentőségű CERT-et, amely valamely jogszabály alapján meg is van említve. Az észszerűségre alapozva a hálózati ábrán mind az öt szervezet össze van kötve egymással, mivel jó eséllyel az egy országban működő eseménykezelő központok, ha nem is minden adatot osztanak meg egymással, de a fontosabb esetekben azért biztos kommunikálnak egymással. A narancssárga entitások azokat a CERT-közösségeket jelzik, amelyek valamilyen internetes forrás alapján köthetők a GovCERT-hez. A lista valószínűleg nem teljes, és további pontok és élek határozhatók meg a hálózatban, azonban a nyíltan elérhető információk nem teszik lehetővé egy mélyebb elemzés elkészítését.



4. ábra

A magyarországi eseménykezelő központok hálózata

Forrás: a szerzők szerkesztése

A következő hálózati ábra azt mutatja be, hogy a négy CERT-közösségben részt vevő CERT-ek, CSIRT-ek hogyan kapcsolódnak össze, azaz milyen lehetséges információáramlási útvonalak léteznek a 2018. november 25-ei, internetes forrásokon alapuló adatok szerint. A hálózati ábra azonban nem teljes, és nem képes teljesen tükrözni a valóságot. Ennek két fő oka van. Ugyan négyből három közösségről hivatalos oldalon közölt adatok találhatók meg, de az IWWN tagjainak forrása nem tekinthető hitelesnek. Az adatok a fent említett dátumnak megfelelően a következő oldalakról származnak:

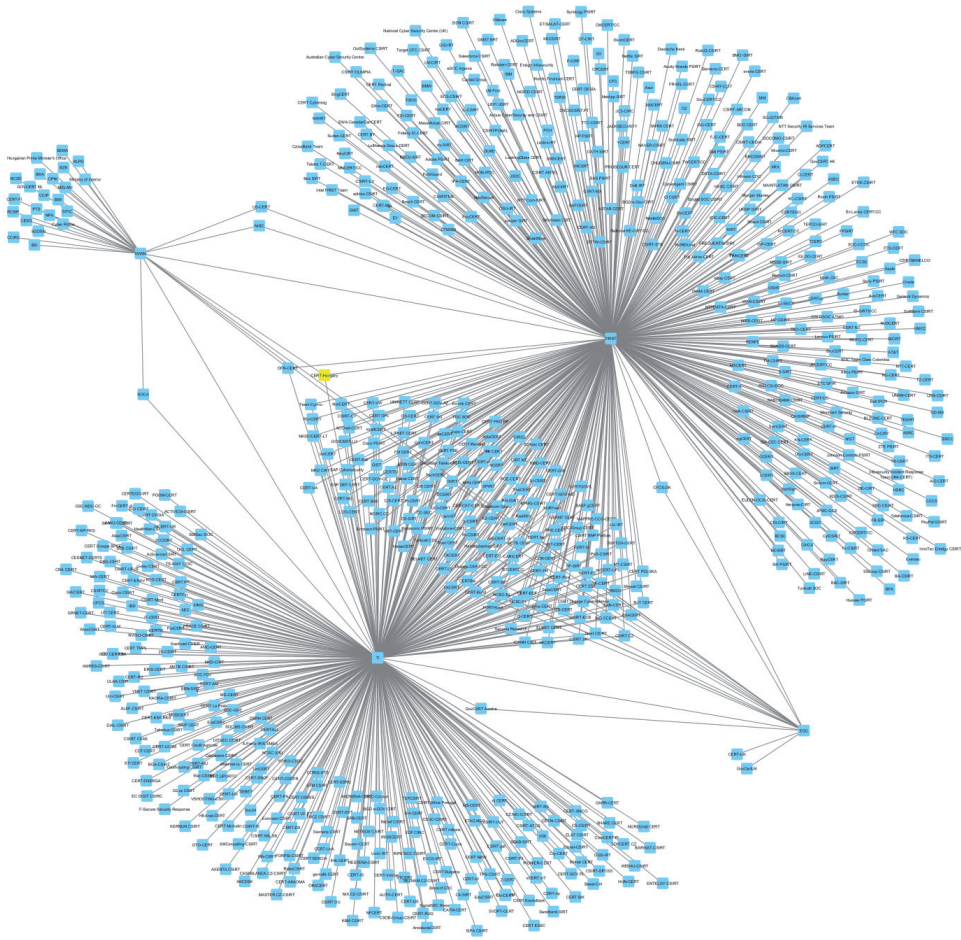
- Forum of Incident Response and Security Teams (FIRST)[7]
- Trusted Introducer (TI)[8]
- European Government CERTs Group (EGC)[9]
- Internet Watch and Warning Network (IWWN)[10]

A teljes hitelességen kívül azonban van más tényező is, ami miatt nem jelenthető ki fenntartás nélkül, hogy a valóságot tükrözi az ábra. Ez pedig a különböző szervezetek nevének az egységes írása. Az adatok elemzése közben kiderült, hogy nem minden csoportnál írták ugyanúgy a szervezetek hivatalos nevét. Erre példa a sárgával jelzett CERT-Hungary is, amely két különböző néven került be a nyilvántartásokba.

Az ábra célja, hogy szemléltesse azt a majdnem 900 kapcsolatból áll hálózatot, amely csupán egy kis szelete a valóságnak, azonban már ebből is jól látszik, hogy alapvető érdeke az eseménykezelő központoknak, hogy részt vegyenek a különböző incidenseket érintő információáramlásban. A négy közösséget a képzeletbeli X végpontjainál található nagyobb csomópontok jelzik. Bal felül az IWWN, jobb felül a FIRST, bal alul a TI és jobb alul az EGC található. Jól látszik, hogy vannak olyan CERT-ek, illetve CSIRT-ek, amelyek több közösséghez tartoznak, így maguk is csomóponttá válnak.

A valóság leképezéséhez azonban ennél jóval több adatra lenne szükség, amelyek nem állnak rendelkezésre jelenleg. Egy részük pedig kifejezetten bizalmas információnak minősül, így a teljes képet talán sohasem sikerül felrajzolni. Ennek ellenére fontos következtetések vonhatóak le. Hálózatelemzési szempontból fontos látni, hogy az így kialakuló több magas fokszámú pont megakadályozza azt, hogy egy-egy szervezet elérhetetlenné válása esetén leálljon a szükséges információk áramlása. Természetesen ezek célzott támadása esetén jelentős problémák jelentkeznének, azonban a hálózatból véletlen elemet kiemelve nagy valószínűséggel nem történne nagy kár.

[7] www.first.org/members/map
[8] www.trusted-introducer.org/directory/country_LICSA.html
[9] www.egc-group.org/contact.html
[10] http://itlaw.wikia.com/wiki/International_Watch_and_Warning_Network



5. ábra

A CERT-közösségek közötti kapcsolat

Forrás: a szerzők szerkesztése

Összefoglalás

A hálózatok elemzésének lehetősége az élet számos területén vezethet olyan összefüggések felismeréséhez, ami alapján értelmezhetővé válik a körülöttünk lévő világ. Jó példa erre a különböző eseménykezelő központok (CERT-ek, CSIRT-ek) ábrázolása. A tanulmányban vett példák nem teljes körűek, és messzemenően nem modellezik pontosan a valóságot, azonban rávilágítanak arra, hogy foglalkozni kell a különböző, e szervezeteket összefogó közösségekkel. A kiberbiztonsági incidensek kezelése alkalmával e közösségek még akkor

is fontos szerepet töltenek be, ha nem nevezhető mindent átfogónak az egymással megosztott adatok köre.

Felhasznált irodalom

ENISA (2016): *Részletes leírás a CSIRT-csoportok létrehozásáról.* Elérhető: www.enisa.europa.eu/publications/csirt-setting-up-guide-in-hungarian/at_download/fullReport (A letöltés dátuma: 2019. 03. 12.)

Haig Zsolt – Várhegyi István (2008): A cybertér és cyberhadviselés értelmezése. *Hadtudomány,* elektronikus szám, 1–12.

Microsoft (2018): *Microsoft Advanced Threat Analytics.* Elérhető: www.microsoft.com/en-us/cloud-platform/advanced-threat-analytics (A letöltés dátuma: 2019. 03. 12.)

NEIH (2018a): *Nemzetközi IT-biztonsági sajtószemle, 2018. 1–2. hét.* Elérhető: http://neih.gov.hu/sites/default/files/dlc/Sajt%C3%B3szemle_1-2.h%C3%A9t.pdf (A letöltés dátuma: 2019. 03. 12.)

NEIH (2018b): *Nemzetközi IT-biztonsági sajtószemle, 2018. 41. hét.* Elérhető: http://neih.gov.hu/sites/default/files/dlc/Sajt%C3%B3szemle_41.h%C3%A9t_0.pdf (A letöltés dátuma: 2019. 03. 12.)

NIS (2016): Az Európai Parlament és a Tanács (EU) 2016/1148 irányelve (2016. július 6.) a hálózati és információs rendszerek biztonságának az egész Unióban egységesen magas szintjét biztosító intézkedésekről. OJ L 194, 19.7.2016.

Tikos Anita (2018): Biztonsági eseménykezelés a nemzetközi térben. A CERT/CSIRT működése. In Berzsenyi Dániel – Gyaraki Réka – Hámornik Balázs Péter – Hirsch Gábor – Kiss Attila – Marsi Tamás – Orbók Ákos – Simon Béla – Solymos Ákos – Tikos Anita – Zsíros Péter: *Incidensmenedzsment.* Budapest–Pécs, Dialóg Campus Kiadó.

Verizon (2018): *2018 Data Breach Investigations Report, Executive summary.* Elérhető: www.verizonenterprise.com/resources/reports/rp_DBIR_2018_Report_execsummary_en_xg.pdf (A letöltés dátuma: 2019. 03. 12.)

Bederna Zsolt – Váczi Dániel – Pollner Péter – Szádeczky Tamás

Támadás hálózatba szervezve

Bevezetés

Az elvárt biztonsági szint eléréséhez vezető úton, valamint a biztonságosnak ítélt úton történő maradás végett döntések sorozatát szükséges meghozni a korlátos erőforrások felhasználási céljáról, módjáról, helyéről és idejéről, mialatt a biztonságot szem előtt tartva a kockázatok útvesztőjében tartózkodunk. Az információbiztonság sikeres megvalósításához elengedhetetlen a támadó fél ismerete.

Az információbiztonság tekintetében a támadó egy olyan valós személyt vagy mesterséges intelligenciát takaró entitás vagy ezek egy csoportja, amely jogosulatlan információszerzés és/vagy a célrendszer által nyújtott szolgáltatások elérhetőségének részleges vagy teljes megakadályozása céljából az adott rendszer biztonságára tör.

A McAfee IT biztonsági szakembere, Robert Siciliano által korábban alkalmazott besorolás alapján a kibertérbeli támadókat a következőképp csoportosíthatjuk: *1. white hat hackers, 2. black hat hackers, 3. script kiddies, 4. hacktivists, 5. state sponsored hackers, 6. spy hackers, 7. cyber terrorists.*[1]

A Trend Micro 2012-ben publikált tanulmánya szerint a támadók számára a feketepiacon elérhető, változatos lehetőségek biztosítottak:

- különböző célra szolgáló, valamint eltérő adottságokkal és biztonsági garanciákkal rendelkező szerver bérlése (0,5–2000 US$);
- DDoS-támadás kivitelezése akár 1 nap (30–70 US$), akár 1 hónap (1200 US$) időintervallumra vonatkozóan;
- spamküldés e-mail, SMS, ICQ, Skype stb. célpontok számára (például a „cheap email services" szolgáltatás 1 000 000 e-mail küldését 10 US$ ellenértékért valósította meg);
- botnet-infrastruktúra bérlése (2000 bot esetén 200 US$).[2]

Ha megvizsgáljuk az internet- vagy a felhőszolgáltatások felépítését, azok a skálafüggetlen hálózatokhoz hasonló jellemvonásokat mutatnak.[3] Egy skálafüggetlen hálózat hátránya a több kapcsolattal rendelkező, tehát a meghatározó pontok elleni célzott támadások esetén mutatkozik meg.[4] Ha ugyanis ezek a pontok kiiktatásra kerülnek, az eredeti hálózat olyan halmazokra esik, amelyeknek nincsen közös elemük – megfosztva azt a feladatának

[1] Siciliano 2011.
[2] Goncharov 2012.
[3] Barabási–Bonabea 2003.
[4] Barabási 2001.

maradéktalan ellátásától. A gócpontokat az internet esetében a nagy hálózati elosztók, míg a felhőszolgáltatások esetén – a Microsoft Azure kiépítettségét alapul véve – néhány meghatározó adatközpont jelentheti.

Ugyanakkor a célzott támadások nemcsak az internet alapjait vagy különböző publikus szolgáltatásokat érinthetik, hanem a kritikus nemzeti infrastruktúrákat (például energiaellátás, katonai infrastruktúra) is. A kibertérben az informatikai rendszerek, szolgáltatások vagy épp kritikus infrastruktúrák elleni támadások jelentős részét különböző, hálózatba kapcsolt, megfertőzött végpontok végzik. Az általuk rendelkezésre álló erőforrást elosztott túlterheléses támadásra (Distributed Denial of Service, DDoS), kéretlen levelek (spam) küldésére és azon túlmenően bármely más irányított támadás, azaz illegális tevékenység céljából bevetheti az irányítója. Másfelől, amennyiben a botnetek felépítésében is azonosítható a skálafüggetlen hálózatokra jellemző felépülés, úgy e jellemző a botnetek kiiktatásában is felhasználható.

A botnetek által megvalósított funkcionalitást és támadásokat figyelembe véve az ENISA[5] a botneteket az egyik legveszélyesebb fenyegetések közé sorolja.[6]

Célkitűzés

A kutatás célja a hálózatokat jellemző sérülékenységek általános jellegű bemutatása, valamint a botnetek mint hálózatok megismertetése a hallgatókkal. A kutatás végrehajtása során elemezzük és bemutatjuk a botnetek jellemzőit, alkotóelemeit, emellett lehetőségeket keresünk az ellehetetlenítésükre, teljes eliminálásukra vonatkozóan.

A tanulmány felépítése

A továbbiakban a tanulmányban a botnetek általános felépítését ismertetjük *(Általános botnetismertető),* ezt követően az általános ismertetőben tárgyaltakra támaszkodva a botnetek fejlődését mutatjuk be *(Botnetevolúció).* A botnetek fejlődése a védelmi képességek fejlődését igényelte, amelyet az általános jellegű képességek és a célzott védelmi megoldások vonatkozásában tárgyalunk *(A detektálási képességek fejlődése),* végül egy tűzfal naplóadatai alapján mutatjuk be a botnetek felderítését *(Botnetgyanús esetek keresése egy tűzfallog alapján).*

Általános botnetismertető

Az esetek többségében a rendszereket ért támadások esetében a támadók által kiaknázható támadási vektorokat az adott rendszer és a felhasználók interakciós pontjai egyértelműen meghatározzák. A támadási vektor olyan attrakciós pontot jelöl, amellyel egy támadó kompromittálhatja az adott informatikai szolgáltatást vagy informatikai rendszert egy

[5] Lásd az ENISA honlapját: www.enisa.europa.eu/

[6] ENISA: *Botnets* (é. n.).

meglévő, ismert sérülékenység vagy egy még ismeretlen módszer által, ezzel lehetőséget adva a rendszerhez tartozó erőforrással történő visszaélésre.

A támadó által használt infrastruktúrák bonyolultsága széles skálán mozoghat – kezdve az egygépes manuális rendszertől egészen a több száz vagy épp több ezer automatikus rendszerig és azokon túl. A botnet az interneten elérhető számítógépek azon csoportja, amelyek erőforrásait az eszköz tulajdonosának és/vagy használójának tudta és akarata ellenére egy támadó vagy támadói csoport ártó szándékkal használja nem legális tevékenység folytatására, azaz egy vagy többféle kibertámadás kivitelezésére.

A támadások motivációja lehet politikai, vallási, haszonra törő, illetve valamilyen egyéb előny szerzésére irányulhat (például gazdasági). Az előnyszerzés akár vállalati, akár állami szinten is értelmezhető, amelyhez a felhőszolgáltatások mintájára elérhetők a kártékony megoldások, eszközök és támadást kivitelező infrastruktúrák. A „szolgáltatási" modellt a *cybercrime as a service*[7] megnevezés takarja, amely azt jelenti, hogy a kibertámadások alapjául szolgáló megoldásokat, infrastruktúrát stb. szolgáltatásalapú üzleti modellre építve kínálják az ügyfeleknek.

A *cybercrime as a service* jelentősebb alkategóriái:[8]

- *Crimeware as a service:* Ezzel a szolgáltatással a sérülékenységek azonosítása és az azt kihasználó, valamint az azt alkalmazó szándékának eleget tevő exploit létrehozása valósítható meg. Jelentős alkategóriát képviselnek a malware-ek, azon belül a nem közismert sérülékenységek, APT-k, rootkitek és a ransomware-ek. A megvalósításukban különböző támogatóeszközök (úgymint dropper, keylogger, bot stb.), rejtőzködést megvalósító megoldások (úgymint cryptor, polimorfizmus stb.) jelentik az építőelemeket. A felsorolt elemeken túlmenően a fizikai eszközök töréséhez szükséges hardvereknek is nagy szerep juthat.
- *Cybercrime infrastructure as a service:* Infrastrukturális elemek, úgymint kliensek és szerverek tartoznak a kategóriába. A kliensek által DDoS-támadás hajtható végre, míg a szerverek a sérülékeny tartalmat (például a weboldalt) teszik elérhetővé. A botnetek ebbe a kategóriába tartoznak.
- *Hacking as a service:* Az egész támadási folyamatot kiszervezik, azt a szolgáltató fél tervezi meg és hajtja végre a megrendelő szándékának megfelelően. Ez lehet akár a működési folyamat akadályozása, akár érzékeny információ eltulajdonítása.

Ha egy eszközre már feltelepült a kártékony kód, az a botnet részévé teszi az adott eszközt (asszimilálja). Ennek következtében az eszköz erőforrásai (számítási, tárolási, hálózati stb. kapacitása) a botnet részévé válnak. További probléma, hogy a lokális és a hálózaton tárolt adatokhoz, valamint akár az eszköz által kezelt felhasználói interakciókhoz is hozzáfér a támadó.

A fentiek értelmében a lehetséges képességek közül a legfontosabb a kémkedés, a spamküldés, a számítási kapacitás kiaknázása „brute force" támadás vagy kriptovaluta bányászása céljából, valamint DDoS-támadás megvalósítása. Ez utóbbival kapcsolatosan az Akamai 2015-ben tette közzé elemzésének eredményét, miszerint a XOR DDoS botnet 150+ Gbps aggregált sávszélességgel rendelkezett a túlterheléses támadások kivitelezésére.[9]

---

[7] Cybercrime as a Service 2013.

[8] McAfee 2013.

[9] Akamai 2015.

Botnetarchitektúra

A botnetek struktúrájukban, képességeikben és technikai megvalósításban is eltéréseket mutatnak, azonban általánosan elmondható, hogy tartalmaznak egy irányító felet (botmaster vagy botherder), egy vagy több irányító szervert (Command and Control, C&C) és egy vagy több irányított gépet (bot).

A botmaster a teljes botnetet vagy annak egy valós részhalmazát irányító entitás (támadó). A botmaster által kiadott utasításokat a C&C-szerver juttatja el a botok számára, betöltve a botnet operatív irányításának feladatát. A bot összességében a fertőzött eszközön futó szoftver, amely ügynök szoftverfunkciót tölt be (agent), végrehajtva a C&C-szervertől kapott utasításokat. Az utasítások végrehajtása mellett támadói szempontból fontos kitétel a botmaster rejtőzködésének biztosítása.

A C&C-szerver elérésének, elérhetőségének módja szabja meg a botnet felépítését. Ennek függvényében centralizált (hierarchikus felépítésű), decentralizált (peer-to-peer), valamint hibrid botneteket különböztethetünk meg.

1. táblázat

Botnetarchitektúra

Típus	**C&C-szerverek száma**
Centralizált	Centralizált felépítés esetén egy vagy több fix számú szerver tölti be a C&C-funkciót, amelyeket (több szerver esetén) hierarchiába is szervezhetnek. A rétegek száma a botnet bonyolultságát és a botmaster elrejtésének valószínűségét növeli, míg a rétegenkénti C&C-szerverek számossága egyrészt terheléselosztást, másrészt redundáns kialakítást szolgál.
Decentralizált	Nincs megkülönböztetett C&C-szerver, mindegyik bot betölti (betöltheti) ezt a funkciót.
Hibrid	Néhány C&C-szerver, amelyek peer-to-peer módon működnek egymás közt.

Forrás: a szerzők szerkesztése

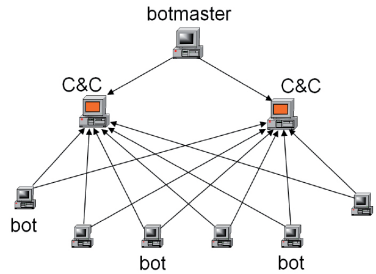
A botnetek felépítésének ismerete a célzott védelmi képességet is hatékonyabbá tudják tenni. Az *A Taxonomy of Botnet Structures* című tanulmányban a kutatók az 1. Erdős–Rényi-modell, 2. a Watts–Strogatz-modell, 3. a Barabási Albert-László-féle skálafüggetlen hálózati modell és 4. a P2P-modell szerint vizsgálták a botnetek felépítését. A vonatkozó kutatások alátámasztották, hogy egyes botnetek skálafüggetlen hálózati modellel jellemezhetők.[10]

Centralizált felépítésű botnetek

A centralizált felépítésű botnetek esetében statikus jellegű a botok vezérlése, azaz egy vagy több előre meghatározott számú elérési móddal rendelkező, fix C&C-szerver található a hálózatban. A C&C elérése IP-cím alapján vagy DNS-rekord lekérdezésével valósul meg. A jellemző kliensszerveri kommunikációs protokoll a HTTP- vagy épp az IRC-protokoll. Az utasítások átadása push és pull módon egyaránt lehetséges. A centralizált botnetek esetében a botmaster számára magas fenyegetettséget jelent a C&C-szerverek viszonylagosan

[10] DAGON et al. 2007.

könnyű azonosítása és azok elérhetőségének lekapcsolása, azaz kiiktatása. E kategóriába tartozik például az AgoBot, RBot és a Zeus botnet.

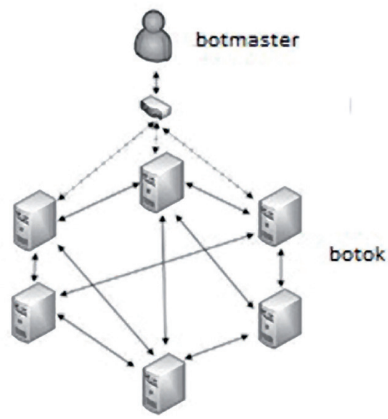


1. ábra
A centralizált botnet architektúrája

Forrás: Wang–Sparks–Zou 2010

Decentralizált felépítésű modellek

A decentralizált felépítésű botnetek vezérlésében nem azonosítható megkülönböztetett C&C-szerver. Ehelyett mindegyik bot nyilvántartja a hálózatban elérhető botok számát, peer-to-peer (P2P) lekérdezési mechanizmussal feltérképezi a környezetében jelen lévő botokat, és lekérdezi az utasításokat (pull), vagy épp – a botmastertől vagy egy másik bottól – megkapott instrukciót adja tovább (push).



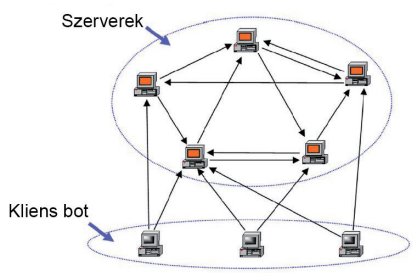
2. ábra
A decentralizált botnet architektúrája

Forrás: Hyslip–Pittman 2015, 12.

A decentralizált botnetek esetén magas fenyegetettséget jelent a botmaster számára a P2P-működés kiépítésének és fenntartásának nagy fokú bonyolultsága, ezzel összefüggésben a botnet dinamikus hálózati felépítés jellegét (dynamic routing) megzavarva a rendszer részeire particionálódhat, adott esetben működésképtelenné válhat. E kategóriába tartozik például a Storm worm, a Nugache és a Conficker botnet.

Hibrid felépítésű modellek

A hibrid botnetek kialakításában a támadók igyekeztek a centralizált és a decentralizált architektúra előnyeit egyesíteni. A C&C-szerverek esetén a felépítés decentralizált jellegű, azaz P2P. Ezzel szemben a botok a C&C-P2P-hálózat egy-egy elméhez csatlakoznak, azaz egy bot egy C&C-szerverhez csatlakozik. Ezzel drasztikusan csökkenthető a teljes botnet-hálózat detektálásának esélye: amennyiben egy C&C-szerver azonosításra és eliminálásra kerül, az a botnethálózatnak mindössze egy valós részhalmazát érinti.



3. ábra
A hibrid botnet architektúrája

Forrás: WANG–SPARKS–ZOU 2010

Működési életciklus

A botnetek működési életciklusa öt fázisra bontható:[11]

1. Kezdeti fertőzés: Az első fázis során a botmaster legalább egy valós sérülékenységet kihasználva, egy vagy több támadási vektoron fertőzi az elérhető sérülékeny eszközöket.
2. Másodlagos injekció: Sikeres fertőzés esetén a bejuttatott kód letölti az ügynökalkalmazást.
3. Kapcsolódás: Az ügynökalkalmazás felveszi a kapcsolatot a botnet C&C-szerverével, amely az eszközt a botnet részévé asszimilálja, az is bottá válik.

[11] FEILY–SHAHRESTANI–RAMADASS 2009.

4. Command and Control: A botok irányítása a botmaster által a C&C-szerveren keresztül kiadott parancsokkal, ami gyakorta egyedi támadások vagy támadási kampányok kivitelezésében testesül meg.
5. Frissítés és karbantartás: Javítások, frissítések, új képességek telepítése, valamint az adott botnetszegmens időszakos vagy végleges lekapcsolása.

Hálózati protokollok

Az IBM X-Force® Research által 2016-ban készített tanulmány szerint a négy leggyakrabban alkalmazott kontrollprotokoll az IRC, a HTTP, a P2P és a Tor.[12] A hálózati kommunikáció vonatkozásában a megoldásokat árnyalja a kriptográfiai megoldások (titkosítás, szteganográfia) alkalmazása, valamint további variánsok megjelenése, illetve alternatív protokollok (például MSN) bevetése.

A rejtőzködés elősegítése végett a botok a C&C-infrastruktúra dinamikus kezelésével veszik fel a szerverekkel a kapcsolatot. Ennek megfelelően a statikus jellegű, IP-reputációalapú vagy hasonló blacklist védelmi képességek nem működőképesek.

A támadók számára az egyik lehetséges technika a fast flux alkalmazása.[13] Továbbá a hálózati protokollok által adott képességeket, azaz a TCP/IP modell szerint az internet, a szállítási és az alkalmazásréteg működési paramétereit, valamint azok esetleges kriptográfiai kiterjesztéseit (például HTTPS, DNSSEC) kiaknázva a detektálási és ennek következtében a védelmi képességek hatékonysága csökkenthető.[14]

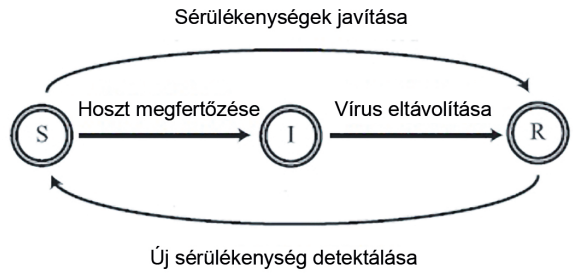
Képességek

A botnetek egy vagy több képességgel rendelkeznek, amelyek közül a fontosabb képességek az alábbiak:[15]

- A botnetek többsége rendelkezik a DDoS-támadás képességével, amely során a támadó leköti a támadott fél erőforrásait, megakadályozva, hogy valós klienskéréseket szolgáljon ki az adott hálózat, szerver vagy szolgáltatás.
- A spyware funkció a fertőzött eszközökön kezelt (például tárolt, megadott) felhasználói adatokat (például hozzáférési adatok, hitelkártyaadatok) küldi meg a botmaster számára.
- A spam képességgel a botmaster kéretlen leveleket képes kiküldeni valós címzettek számára.
- Fraud, azaz csalás kivitelezésekor a fertőzött eszközöket például egy website felkeresésére utasítják.
- Kriptovaluta-bányászat.
- Önterjesztési képesség.

[12] IBM Corporation 2016.
[13] IBM Corporation 2016, 14.
[14] Acaralia et al. 2016.
[15] IBM Corporation 2016, 15–18.

A felsorolásban nem szerepel, de a botmaster szempontjából kiváltképp fontos az önfenntartás képessége, amely tulajdonképp összefügg az önterjesztési képességgel. A *Stochastic modeling of self-evolving botnets with vulnerability discovery* című tanulmány szerzői egy négyállapotú (Susceptible-Infected-Recovered-Susceptible) SIRS-modellel jellemezték a botnetekre jellemző állapotokat.[16] Az állapotátmeneteket az alábbi ábra szerint lehetséges lefolytatni:



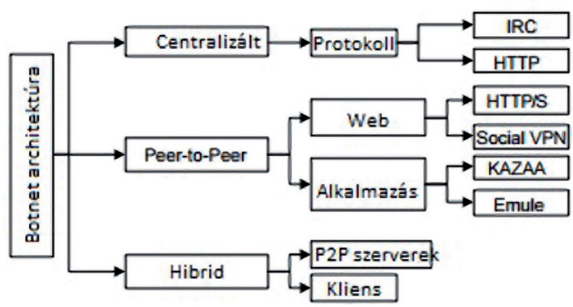
4. ábra
SIRS-modell

Forrás: KUDO et al. 2018, 103.

A tanulmány készítői arra a következtetésre jutottak, hogy az önfejlődésű botnetek autonóm sérülékenységfelfedező és -kihasználó képességekkel rendelkeznek, amelyekkel biztosítható a dinamikus működésük.[17]

A botnetek csoportosítása

A botneteket számos paraméter szerint csoportosítjuk. Az alapvető megkülönböztető jegyek: az architektúra és a protokollok, ahogy azt az alábbi ábra is mutatja:



5. ábra
A botnetek csoportosítása

Forrás: KARIM et al. 2014

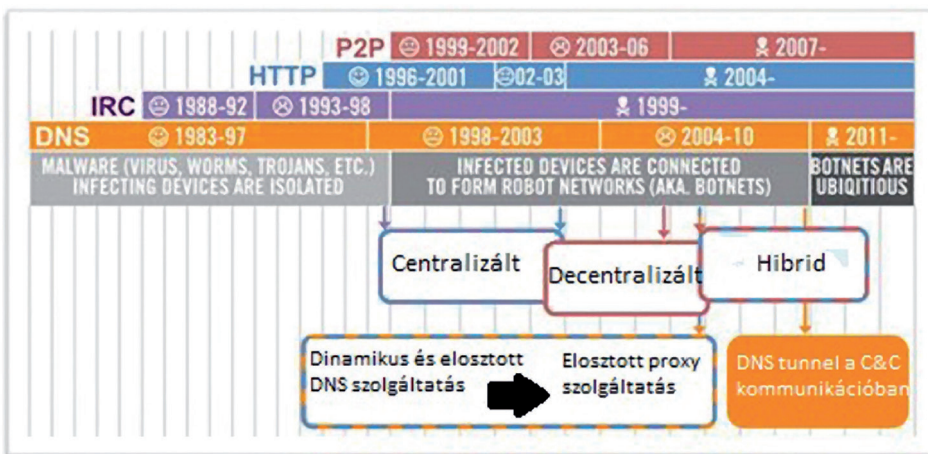
[16] KUDO et al. 2018.
[17] KUDO et al. 2018, 109.

További meghatározó jellemzőnek számítanak a botnetek által implementált képességek. Azonban a működési környezetet megvizsgálva további paraméterként adódhat 1. a hatásosság (effectiveness), 2. az átlagosan elérhető sávszélesség (average available bandwidth), 3. a hatékonyság (efficiency) és 4. a robusztusság (robustness).[18]

Botnetevolúció

Történelem

Az első internetes férget Robert Morris Jr., a Cornell University végzős hallgatója készítette még 1988-ban, amely a „hazatelefonálás" képességével rendelkezett a Berkeley által üzemeltetett C&C-szerver felé. Ezt a megoldást folyamatos újítások követték.[19] Az alábbi ábrának megfelelően a tényleges botnetek története 1999-ben kezdődött a Sub7 trójaival és a Pretty Park féreggel, amelyek centralizált architektúrájú felépítésben IRC-protokollt alkalmaztak. 2000-ben a mIRC-kliens „továbbfejlesztéseként" a GTbot indult útjára, majd 2002-ben az SDBot már C++ alapon íródott. Az SDbot forráskódját a készítője értékesítette. Az Agobot új megközelítésként moduláris, azaz többfázisú támadást alkalmazott.



6. ábra
Botnetevolúció

Forrás: Cantón 2015

2003-ban az SDBot botnetből a Spybot botnet született, amellyel új képességeket is kapott a hálózat, úgymint bevitt karakterek rögzítése, adatbányászat, valamint az azonnali üzenetküldő rendszereken keresztül megvalósított kéretlen üzenetek küldése (spammed instant

[18] Dagon et al. 2007.
[19] Ferguson 2010.

messages, spim). Az Rbot a DDoS- és adatlopási képessége mellett a kommunikációban SOCKS-proxyt használt, mindamellett elsőként tömörítési és titkosítási algoritmust alkalmazott a detektálás elkerülése céljából.

2004-ben a Polybot elsőként alkalmazott polimorf[20] algoritmust kódja megváltoztatására, továbbá a botnetek az IRC-C&C-csatornáról végleg váltottak a http-, ICMP-protokollokra, továbbá adott esetben már SSL-alapú, titkosított csatornát is használtak.

Az időközben már többszörös verzióváltáson és frissítésen átesett Zeus crimeware család működése 2006-ban adatlopási képességgel indult útjára. A frissítések során gyakorta új képességeket is kapott a hálózat. Az új képességgel ellátott botnet egy valós részhalmazát bérleményként igénybe is lehet venni.

A kezdeti megoldást, a C&C-szerverek IP-címének statikus jellegű kezelését a biztonsági kutatók egyre magabiztosabban kiszűrték. Erre a Conficker a Cutwail által már korábban alkalmazott napi szintű C&C-szerver alternatív nevek generálását vetette be. A Conficker naponta 50 000 alternatív nevet generált.

Aktuális trend

Idővel a botmasterek legfőbb céljává a pénzügyi haszonszerzés vált, amely motiválta a kártyaadatok, a bankszámlák kezelésével kapcsolatos adatok eltulajdonítását, valamint az illegálisan (az erőforrás tulajdonosa tudta és beleegyezése nélkül) működtetett kriptovaluta-bányászatot. Ezt támasztja alá, hogy a 2011-ben létrehozott Ramnit botnet volt az első bankolási adatok eltulajdonítására specializálódott botnet, amelyet sikeres deaktiválását követően idővel újból aktiváltak.[21]

A mobilbotnetek története is ebben az időben indult útjára a Gemini (2010), valamint az AnserverBot (2011) botnettel. A mobilbotok esetében a botmasternek kényesen ügyelnie kell az eszköz erőforrásainak használatára abból az okból kifolyólag, hogy a „túlzott" számítási kapacitást igénylő műveletek, valamint a magasabb hálózati forgalom az akkumulátort könnyűszerrel lemerítik, és ez felhívja a használójának figyelmét rá. További problémát jelenthet a botok és a C&C-szerver közötti kapcsolattartás.[22]

A social network lehetőségeit kiaknázva a socialbotok a C&C-szerverekkel egy független, magas rendelkezésre állású szolgáltatáson kriptográfiai megoldásokat (például titkosítás vagy szteganográfia) latba vetve képesek a kapcsolatot kezdeményezni és fenntartani.[23]

Az IBM Corporation által 2016 márciusában készített és megjelentetett *The inside story on botnets* című írásban már említik az Internet of Things eszközökből létrehozott botnet, azaz a „thingbot" rendszerét. A mindössze egy évvel később (2017 áprilisában) megjelent, *The weaponization of IoT devices* című írásban a thingbotok jelentőségének növekedése

[20] Az előrekódolás helyett különböző mutációs technikák alkalmazásával biztosított a dekódolórész változása, például a „polimorf vírusok végtelen számú új dekódolót hozhatnak létre, amelyek különböző titkosítási módszereket alkalmaznak a vírustörzs konstans részének titkosítására (az adatterületek kivételével)". Bijalwan et al. 2016, 502–504.

[21] IBM Corporation 2016.

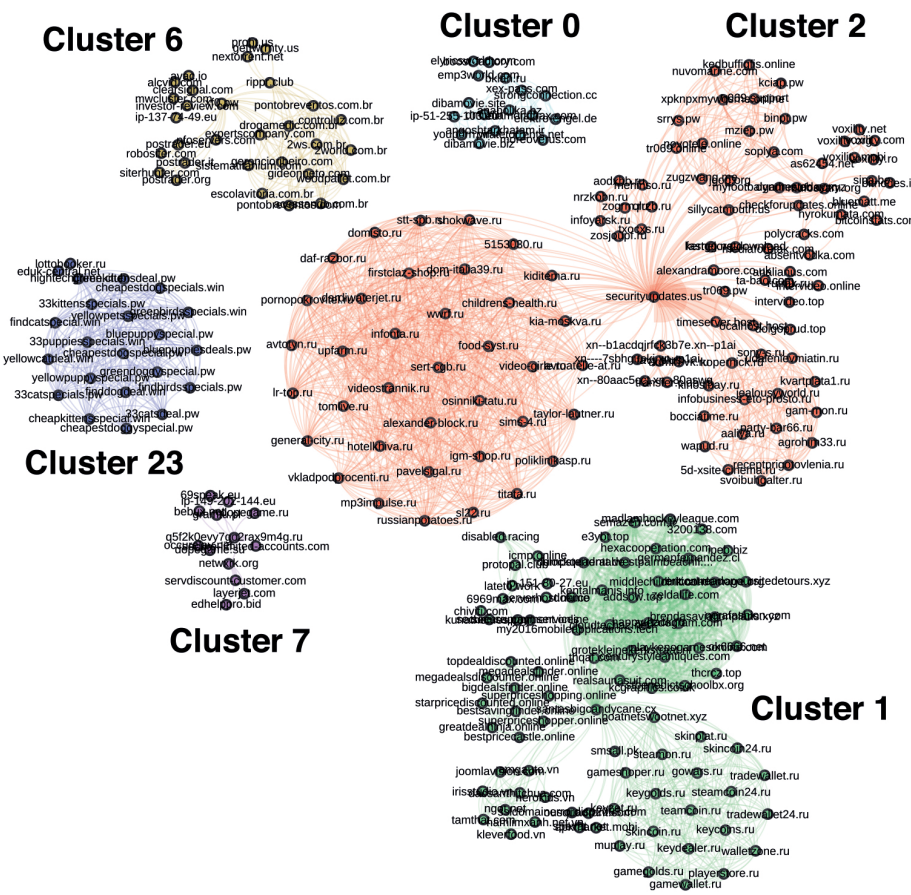
[22] Chen et al. 2017, 270–271.

[23] Boshmaf et al. 2013, 556.

egyértelművé vált. Ennek tanúbizonysága, hogy 2016-ban számos támadást hajtottak végre thingbotokon keresztül, amelyek egyik jelentős képviselője a Mirai botnet volt.[24]

A Mirai botnet felépítése

A Mirai botnet rövid időn belül (kb. 20 óra alatt) mintegy 65 000 IoT eszközt fertőzött meg, így gyakorlatilag ez a mennyiség kezdeti populációként értelmezhető. Továbbá állandó jelleggel 200-300 000 botot irányított, amellyel összességében 15 000 fölötti számban indított támadásokat.[25]



7. ábra

A Mirai botnet és a C&C-klaszterek

Forrás: ANTONAKAKIS et al. 2017, 10.

[24] IBM Corporation 2017.

[25] ANTONAKAKIS et al. 2017.

A Mirai botnet visszafejtésével 33 darab C&C-klasztert azonosítottak, amelyek egymással nem álltak kapcsolatban, azaz különálló infrastruktúrát képeztek. Az információ birtokában feltételezhető, hogy több operátor irányította a botnet működését. Az egyes klaszterek elkülönülő jellege következtében a célzott kiiktatási kísérlet jelenti a leghatékonyabb megoldást (lásd: *A védelmi képességek célzott fejlesztése*).

Detektálási képességek fejlődése

A kiberteret jellemző állandó jellegű harcban az egyik felet a rendszer védelmezői (operátorok, adminisztrátorok, biztonsági szakemberek, kutatók stb.) képezik, míg a másik felet a rendszer támadói jelentik (jelen esetben a botmaster). A szereplők által végrehajtott, a védekezést vagy a támadást megvalósító akciók sorozatának együttesén értelmezhető a játékelmélet,[26] amely szerint egy játékot a játékosok, a játékosok számára lehetséges lépések, valamint az információ struktúrája határoz meg. Az információ, illetve az informatikai biztonság terén való alkalmazásának közös jellemvonása a két játékos számára rendelkezésre álló limitált erőforrás-mennyiség. A biztonsági játék kimenetét a játékosok akciója, reakciója adja meg, amelyet az egyes lépésekhez társított nyereségek (negatív értelemben a költségek) befolyásolnak. Az adott játékos által végzett akciók sorozata adja az adott játékos stratégiáját. E játékot az információbiztonság területén vizsgálhatjuk rövid távon vagy hosszú távon, valamint mikro- és makroszinten egyaránt.

Hosszú távon, makroszinten vizsgálva a botnetek által képviselt fenyegetés kérdéskörét, a botnetek fejlődése (lásd: *Botnetevolúció*) a detektálási képességek fejlődését inspirálta. Mindazonáltal általános érvényű megállapítás, hogy a botnetek detektálása és eliminálása *a priori,* azaz a botnetkommunikáció előzetes ismerete nélkül nagyon nehéz kihívás, minthogy nincs pontos ismeret az egyes végpontok közötti kommunikáció valós vagy kártékony jellegére vonatkozóan.[27]

Általános jellegű védelmi képességek

A botnetek fejlődésével együtt az anti-malware- és az IDPS- (Intrusion Detection and Prevention System) technológiák számos újításon mentek keresztül, ennek következtében manapság a védelmi megoldások számos metodológiát alkalmaznak. Alapvető technikát képvisel 1. a signature-based, 2. a policy-based, 3. a stateful packet inspection, valamint 4. az anomaly-based technológia.[28]

A signature-based megoldás a gyártók által kiadott szignatúrák alapján nagy hatékonysággal képes az olyan fenyegetettségek ellen védelmet nyújtani, amelyet már felismertek, és ellenében szignatúrát is kibocsátottak. Ez azonban a hálózatfelderítés tényének felismerésében vagy a szignatúra-adatbázis frissítésének elmaradása esetén az újonnan detektált fenyegetettségek ellen nem tud segítséget nyújtani.

---

[26] Alpcan–Başar 2011.
[27] Rrushi–Mokhtari–Ghorbani 2011, 792.
[28] Bederna 2015, 149.

A policy-based megoldás a tevékenységek előre beállított policy szerinti megfelelőségének monitorozását valósítja meg. Minden olyan eseményt detektál, amely a policy beállításainak nem tesz eleget, ennélfogva a policy megalkotása magasabb szintű szaktudást és a beállítások finomhangolását igényli.

A stateful packet inspection megoldás olyan profilozáson alapuló (profile-based) metodológia, amely a hálózati protokollok RFC-ben meghatározott szabványától történő eltérését figyeli. Problémát jelent a szabványtól eltérő, gyártóspecifikus protokollimplementációk megléte. További problémát jelent a detektálás ilyen irányú kiterjesztésének erőforrás-igényes jellege.

Az anomaly-based megoldás olyan profilozáson alapuló (profile-based) metodológia, amely akár felhasználói, akár host vagy épp a hálózati kapcsolatok szintjén képes a normálistól történő eltérések detektálására. A problémát a „normális" viselkedési minta meghatározása jelenti.

A stateful packet inspection és az anomaly-based védelmi megoldás esetében a profil meghatározása történhet az időre nézve statikusan, de a nagyobb változtatások alkalmával megismétlődő vagy az idővel dinamikusan változó jelleggel. A dinamikus jelleg alkalmazása esetén nagyobb a valós veszély és a fenyegetettség mértéke.

A botnetek detektálása megvalósítható honeynet, valamint hoszt- vagy hálózati IDS-rendszer alkalmazásával.[29] A védelmi képességek megerősítése végett a mesterséges intelligencia és az annak részét képező gépi tanulás témakörében a neurális hálózatok, korreláció, asszociáció stb. alkalmazása merülhet fel.[30] Ezen túlmenően, a hagyományos kétállapotú logikától eltérően a fuzzy logika alkalmazása is jelentős eredményeket hozhat.[31]

A korreláción alapuló intelligens megoldás a hálózati aktivitás jelentette különböző attribútumok elemzésén alapul. A korrelációs technológia esetén vertikális és horizontális megoldás különböztethető meg. A horizontális korreláció a hosztok viselkedésében, kommunikációjában lévő hasonlóságot vizsgálja, míg a vertikális korreláció az eszközök viselkedését a meglévő botnetviselkedési modellekhez hasonlítja.[32]

A védelmi képességek célzott fejlesztése

Számos kutatás célzottan foglalkozott a botnetek viselkedésével, kialakításával, méretével, képességeivel, az általuk jelentett fenyegetés méretével, valamint a detektálási stratégiákkal, képességekkel és a vonatkozó trendekkel.[33] Egyelőre megkerülhetetlen igazság, hogy „a már létező és működő botnetekre vonatkozó jellemzők meghatározásának fontos és meghatározó szerepe van a hatékony védelmi metodológiák tervezésében".[34]

Alapul véve Dagon és társainak tanulmányában a botnetek felépítésére vonatkozó modellezési vizsgálatot, a felépítés alapján célzott védelmi mechanizmus készíthető. A véletlen

[29] Silva et al. 2013, 387.
[30] Bederna 2015.
[31] A fuzzy logika a kétértelműségből, a pontatlanságból és az információhiányból fakadó bizonytalanság kezelésére alkalmas matematikai apparátus. Bederna 2013, 22–23.
[32] Khattak et al. 2015, 145.
[33] Khattak et al. 2015, 144.
[34] Correia et al. 2012, 165.

felépítésű botnetek magas ellenállással bírnak a behatások kezelésére, amelynek leghatékonyabb kezelési módja a végpontok nagy számosságú kiiktatásának egyszerre történő megvalósítása. Ezzel szemben a skálafüggetlen hálózati modell szerint felépülő botnetek kiiktatására tett célzott műveletek nagy hatékonysággal bírnak.[35]

A botnetekre vonatkozó detektálási képességek összehasonlítása bonyolult és nem egyértelmű feladat, továbbá kevés kutatás foglalkozik e témakörrel. A feladat bonyolultsága a felhasználható adathalmazok, valamint az összehasonlító metodológiák hiányából fakad. Ezt a hatást tovább erősítik a kutatásonként eltérő módon alkalmazott mérőszámok.[36] Mindezek ellenére a detektálási képességek implementálására tett közösségi javaslat alacsony prioritással és hatékonysággal valósult meg, a gyakorlatias tudás az információbiztonsággal foglalkozó gyártóknál keletkezik és marad meg. Az első célzott közösségi védelmi megoldás a Botflex volt.[37]

A korai botnetdetektálási megoldások a payload szignatúra alapú vizsgálat alapján valósultak meg. Azonban a hálózati forgalom vizsgálata a hálózati csomag által hordozott hasznos teher (payload) jellegétől (például titkosítás) függetlenül is működőképes. A gépi tanuláson alapuló klasszifikációt megvalósító REPTree (Reduced Error Pruning) FAR = 0,01%, FNR = 1,70%, valamint DR = 98,30% mutatószámértékek mellett működött, amelyhez az alábbi attribútumokat vizsgálta:[38]

- SrcIp: Flow source IP address (az adott kommunikációs kapcsolatforrás IP-címe),
- SrcPort: Flow source port address (az adott kommunikációs kapcsolatforrás TCP- vagy UDP-portja),
- DstIp: Flow destination IP address (az adott kommunikációs kapcsolat cél-IP-címe),
- DstPort: Flow destination port address (az adott kommunikációs kapcsolat cél-TCP- vagy UDP-portja),
- Protocol: Transport layer protocol or mixed (a szállítási réteg vonatkozó protokollja),
- APL: Average payload packet length for time interval (átlagos hasznos teher mérete adott időintervallumon vizsgálva),
- PV: Variance of payload packet length for time interval (a hasznos teher varianciája adott időintervallumon vizsgálva),
- PX: Number of packets exchanged for time interval (a kommunikáló felek között váltott csomagok számossága adott időintervallumon vizsgálva),
- PPS: Number of packets exchanged per second in time interval T (T időintervallum alatt a kommunikáló felek által váltott csomagok számossága),
- FPS: The size of the first packet in the flow (a kommunikációs kapcsolatban az első csomag mérete),
- TBP: The average time between packets in time interval (a csomagküldések között eltelt átlagos időintervallum nagysága),
- NR: The number of reconnects for a flow (a kommunikációs kapcsolatra vonatkozóan az újracsatlakozások számossága),

[35] Dagon et al. 2007.
[36] García et al. 2014, 101.
[37] Khattak et al. 2015, 144.
[38] Zhao et al. 2013, 6.

- FPH: Number of flows from this address over the total number of flows generated per hour (az adott IP-cím vonatkozásában a kommunikációs kapcsolatok kezdeményezésének aránya egy óra alatt).

A fenti felsorolással ellentétben a botok és a C&C kommunikációjára vonatkozó vizsgálat során az alábbi paraméterek elemzése is megvalósítható:[39]

- a kommunikáció kezdete és vége,
- az alkalmazott protokoll,
- a TCP-kommunikáció során alkalmazott aktív flagek száma,
- a kommunikáció során küldött csomagok száma,
- a kommunikáció során küldött bájtok száma,
- a kommunikáció során eltelt idő mértéke,
- a kommunikáció kezdeményezője (szerver vagy kliens),
- a csomagonként küldött átlagos bájtok számossága,
- a sikeres C&C-kapcsolatok száma,
- a DNS-lekérdezések száma, valamint
- a C&C-szerverrel történő kommunikáció periodicitása.

A hálózati csomagok fejlécén végzett elemzés 1. a Randomized Filtered Classifier, 2. a Logistic Regression, 3. a Random Committee, 4. a Random Subspace és 5. a Multi Class Classifier gépi tanulási klasszifikációs technikákat hasonlította össze,[40] és ennek fényében a Logistic Regression Classifier algoritmus a legalkalmasabb megoldás a botnetek detektálására.[41]

A kommunikáció vizsgálatán alapuló megoldások megfelelő működéséhez elegendő a hálózati forgalom töredékének elemzése. A gyorsított döntési fa (Boosted Decision Tree),[42] a Naive Bayes osztályozó (Naive Bayesion Classifier) és a Support Vector Machine adatbányászati algoritmusokat a kisméretű csomagok (Small_packets), a csomagok aránya (Packet_Ratio), a kezdeti csomag mérete (Initial Packet_length), a bot válaszcsomagja (Bot Response Packet) és a kommunkációs vektor (Flow vector) statisztikai változókon alkalmazva a Naive Bayes osztályozó algoritmus volt a leghatékonyabb a tesztelt megoldások közül.[43]

Ezzel szemben rendelkezésre álló adathalmazokból az ISCX-adathalmazon a kapcsolódó kutatás szerint egy kombinált klasszifikációs algoritmus nagyobb hatékonyságot képes elérni az önálló algoritmusokhoz képest. Az ISCX-adathalmazt 42 attribútum alapján, tanítási és tesztelési részadathalmazra felbontva, egy a k-legközelebbi szomszédon alapuló osztályozás (k-nearest neighbor classifier) és döntésifa-alapú (decision tree) bagging

[39] Correia et al. 2012, 160.
[40] Mathur–Raheja–Ahlawat 2018, 1672.
[41] Mathur–Raheja–Ahlawat 2018, 1676.
[42] A gyorsított döntési fa egy speciális AdaBoost algoritmus.
[43] Kirubavathi–Anitha 2016, 94–95.

(zsákolás), AdaBoost (azaz adatptive boosting, adaptálódó gyorsítás) és softvoting-alapú (szavazás) kombinált algoritmus[44] dolgozta fel.[45]

További hatékonyságnövelő lehetőség a *Detecting botnet by anomalous traffic* című tanulmány készítői szerint, hogy a lehetséges anomáliák vizsgálatát két különböző időtávon javasolt megvalósítani.[46]

Botnetgyanús esetek keresése egy tűzfallog alapján

Nagyobb cégek, intézmények tűzfalat üzemeltetnek, amellyel szervereiket védik. Az alapvető technikai részleteket az *Incidenskezelés a közigazgatásban* című részben ismertettük. Itt a rendszer alapvető topológiáját és néhány konkrét forgalmi tulajdonságot mutatunk be, amelyek támadásra, illetve fertőzésre utalnak.

A tűzfalon átmenő forgalom nagy léptékű áttekintése

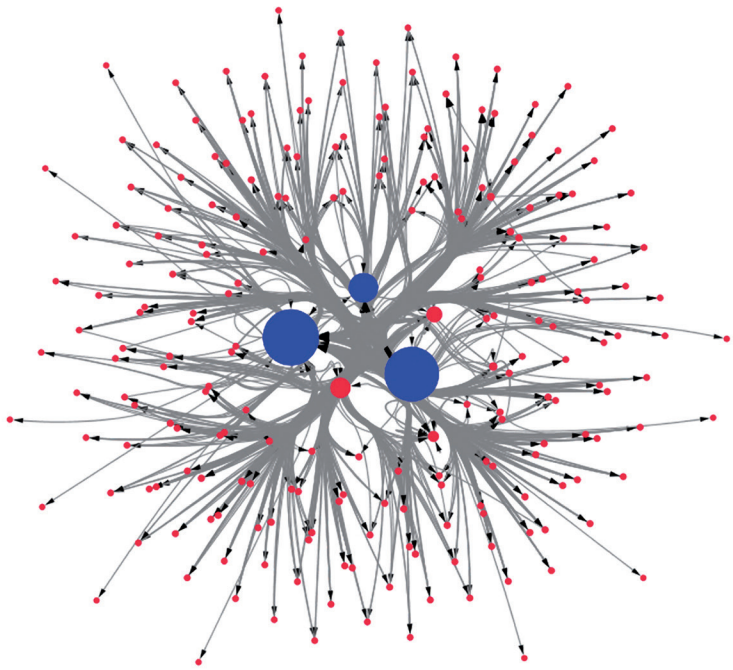
A tűzfal alapvetően két tartományt választ szét: a védendő intézményi szervereket a világ többi részétől különíti el. Az átmenő forgalom elvileg egy páros gráfot kellene, hogy meghatározzon, ha a tűzfal belső és külső oldala egyértelműen lenne szabályozva. Azonban néhány IP-cím mind a védett, mind a külsőnek számító tartományban előfordul, ezért csak közelítőleg érvényesül a páros tulajdonság. Az alábbi ábra szemlélteti az egymással forgalmat bonyolító IP-címek nagy léptékű hálózatát, ahol az IP-címekből csak a legmagasabb szintet, azaz a legelső bájt értékét tüntettük fel. A nem tökéletes párosság ellenére is jól elkülönül a védendő tartomány a külső címek tartományaitól.

A 8. ábrán bemutatott hálózat centrumában található a három védendő tartomány, míg a sugaras alakban a külső címek sorakoznak. Megjegyezzük, hogy a hálózatnak van egy köztes tartománya is, ahol nem magányos végpontok, de nem is a nagy centralitású csúcsok szerepelnek. Ezek az IP-címek azok, amelyek a védendő tartományban vannak, de a forgalmuk teljesen aszimmetrikus: csak adatforgalom kiindulópontjaként szerepelnek, de feléjük nem megy kívülről forgalom. Ezek a címek feltehetőleg valamilyen kiszolgáló intézmény gépei lehetnek, amelyek a lakosság számára nem szolgáltatnak adatot, viszont a munkájukhoz szükség van arra, hogy információt gyűjtsenek az internetről.

[44] Egy adathalmazon végzett osztályozás pontossága több osztályozó előrejelzéseinek kombinálása által megemelhető. E megoldás az együttes (ensemble) vagy kombinált osztályozó (classifier combination) módszer. SZŐR 2010.

[45] BIJALWAN et al. 2016.

[46] CHEN–LIN 2015.



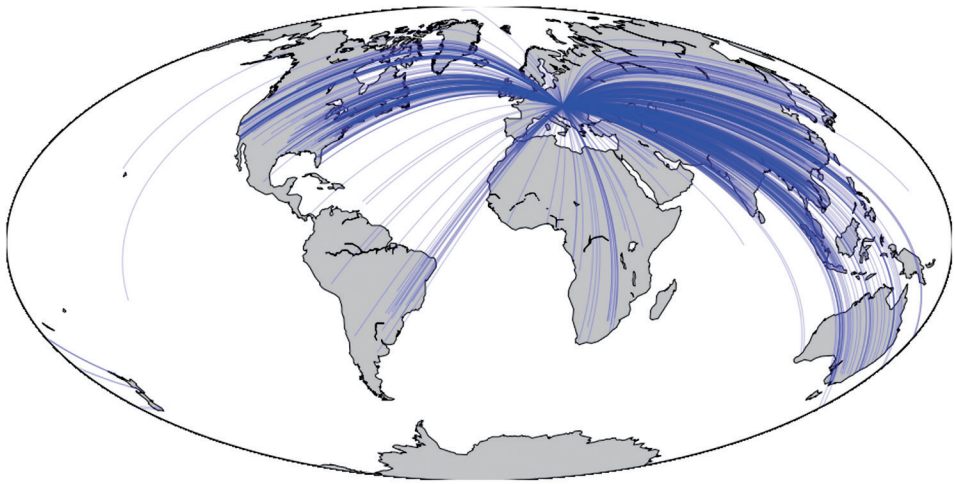
8. ábra

A tűzfalon átmenő forgalom nagy léptékű hálózata

Forrás: a szerzők szerkesztése

A mintaként feldolgozott forgalom IP-címei alapján nemcsak a címtartományok közti áttekintő kép készítése lehetséges, hanem a címeket térképre is lehet helyezni. Természetesen valamennyi különbség van a normál „postai" címekhez képest. Vannak például úgynevezett belső használatú IP-címek, amelyeket csak a tűzfal mögött lehet használni egymás közti kommunikációra, és kifelé a tűzfal ad nekik valamilyen kívülről látható, ideiglenes vagy hosszú távra rögzített címet. Másik eltérő tulajdonság lehet, hogy egyetlen IP-cím forgalmát több gépből álló felhő biztosítja, amelynek a gépei biztonsági szempontok miatt több helyszínre vannak szétszórva. Ezektől a különleges esetektől eltekintve az IP-címek nagy része viszonylag jól geolokálható, legalábbis olyan pontossággal, amellyel a teljes Földön elhelyezhetők egy áttekintés kedvéért.

Az alábbi kép egy ilyen térképes beágyazáson szemlélteti, hogy a feldolgozott forgalomban honnan érkeztek a megkeresések a magyarországi tűzfal mögött elhelyezkedő címekhez.



9. ábra
A tűzfalon átmenő forgalom térképes beágyazása

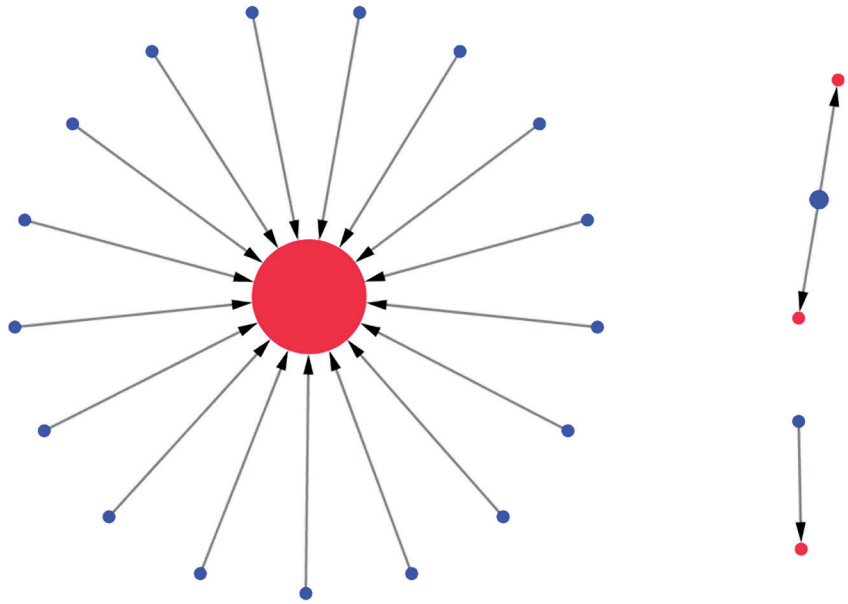
Forrás: a szerzők szerkesztése

SYN FLOOD támadások

A szemléltető példánkban használt adatok az egyik legjobb minőségű és nagy teherbírású céleszközről származnak. Ez az eszköz számos bonyolult mintázatú támadást képes észlelni, megjelölni és elhárítani. Az egyik ilyen, igen komoly károkozásra képes támadási típus a SYN FLOOD támadás. Ennél a támadásnál a külső gép olyan, nagy intenzitású kéréssorozatot küld a kiszolgáló szervereknek, amelyek lefoglalnának minden erőforrást, így a szerver a valódi igényeket már nem tudná teljesíteni.

Az alábbiakban azt vizsgáljuk, hogy azok a gépek, amelyek ilyen típusú támadást intéztek a tűzfal mögötti gépek ellen, milyen potenciális veszélyt jelentenek, illetve hogy csak elszigetelt és semlegesített veszélyforrásról van-e szó.

A vizsgált 14 órányi időszak alatt ezzel a típusú támadással kapcsolatban 195 darab külső IP-címet jelölt meg a tűzfal. A következő hálózati ábra azt mutatja, hogy ezek közül melyek álltak kapcsolatban a védendő tartományban elhelyezkedő gépek valamelyikével.



10. ábra

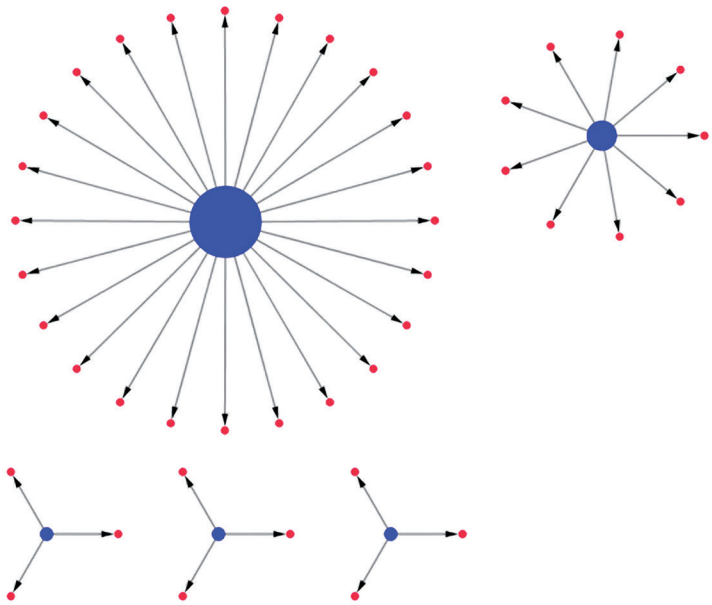
A SYN FLOOD támadókkal kapcsolatot létesítő gépek hálózata

Forrás: a szerzők szerkesztése

A 10. ábrán látható, hogy a 195 támadó gépből csupán 3 olyan van, amivel a védendő gépek kapcsolatba kerültek. Itt a kapcsolat létrehozója nem a támadó gép volt, hanem a tűzfalon belülről kezdeményezték a kapcsolatot. A kapcsolatfelvételek feltehetőleg függetlenek egymástól, mert az egyes komponensek függetlenek, nincs köztük közös kapcsolatfelvételi él. A támadók eltérő sikerességgel tudták magukhoz csalogatni a belső gépek felhasználóit. Két támadóval éppen csak akadt valaki, aki kapcsolatba lépett, míg a harmadik irányába már számos IP-címről folyt kommunikáció.

Megvizsgálható, hogy azok az IP-címek, amelyek a támadók irányában kapcsolatot kezdeményeztek, vajon milyen más címekkel kerültek még kapcsolatba. Azaz ha esetleg a kapcsolatfelvétel alkalmával megfertőződtek, akkor merre várható a fertőzés továbbterjedése. Mivel nagy forgalmú IP-címekről van szó, csak azokra a terjedési lehetőségekre koncentrálunk, amely irányokban a forgalom az átlagostól eltérő jelleget mutatott. Átlagos alatt itt a legnagyobb adatforgalmú vagy a napi működéssel együtt járó szolgáltatások használatát értjük, mint például www, e-mail, login.

Az alábbi ábra a 10. ábrán látható hálózat kiegészítése. Itt az eredeti támadókon és a velük kapcsolatot kialakító, belső címeken kívül azok a kapcsolatok is szerepelnek, amelyek úgynevezett port-scan jellegűek.



11. ábra

A SYN FLOOD támadókkal kapcsolatot létrehozó gépek hálózata

Megjegyzés: kiegészítve azokkal a további forgalmakkal, amelyek a tűzfal belső oldalán lévő gépekből indultak, és a célgépen az átlagosnál több portra kapcsolódtak.

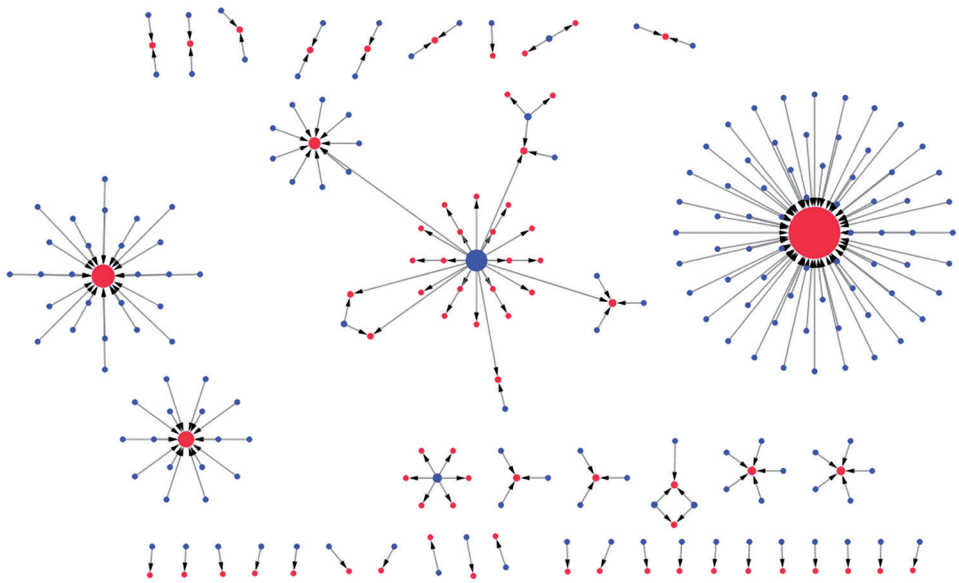
Forrás: a szerzők szerkesztése

Port-scan jellegű támadások

A port-scan jellegű támadás egy tipikusan automatizált támadási forma. Itt a megtámadott gép gyengén védett vagy a külvilág felé nyitott szolgáltatásait teszteli a betörő. Egy gép karbantartása, fejlesztése vagy tesztelése során sokszor előfordul, hogy nyitnak egy új szolgáltatást vagy belépési lehetőséget egy olyan portszámon, amely nem szabványos (például webszolgáltatás a 88888-as porton). Vagy éppen a fordított eset is előfordulhat, azaz egy régóta nem karbantartott gépet újra bekapcsolnak, és a régi szoftverekről idő közben publikussá vált hibákat kihasználva lehet a gépre betörni.

Ahhoz, hogy a támadó megtalálja a célgépen a gyenge szolgáltatást, végig kell próbálgatnia a lehetséges portszámokat. Ezért is nevezik ezt a típusú támadást port-scan jellegűnek. Alkalmasan konfigurált tűzfalakkal, mint akár a vizsgálat tárgyát képező tűzfal esetén is, ezek a típusú támadások viszonylag könnyen észlelhetők, ha a támadónak kevés eszköz és rövid idő áll rendelkezésére. Az észlelést az nehezítheti, ha a támadás mind időben, mind az eszközök szintjén széles tartományt ölel fel. A támadások súlyosságát

aszerint is meg lehet ítélni, hogy egy-egy gépet hány külső gép szkennel, illetve a szkennelést végző gépek csak egy-egy gépet választanak célpontul, vagy tömegesen, az elérhető gépek jelentős részét megpróbálják végigpásztázni. A támadás ilyen jellegének elemzését segítheti az alábbi ábra. Itt azon forgalmi kapcsolatokból készült a hálózat, amelyekben a kapcsolatfelvevő több mint 100 portszámot próbált végig a célgépen.



12. ábra

Port-scan jellegű támadási kapcsolatok hálózata

Forrás: a szerzők szerkesztése

A 12. ábrán bemutatott hálózat alapján megállapíthatjuk, hogy a port-scan támadások egy része véletlenszerű pásztázás eredménye. Ezek az események az egymástól elkülönülő, egyetlen élből álló komponensekként jelennek meg az ábrán: a támadó IP-cím csak egyetlen cél-IP-t választott ki, mégpedig olyat, amellyel rajta kívül senki más nem áll port-scan jellegű kapcsolatban. A hálózatban láthatunk egymáshoz kapcsolódó gráfkomponensekből álló mintázatokat is, azaz néhol több külső gép támad egy belső gépet, másutt egy külső gép egyszerre több belső gépet is támad. Mivel egy-egy belső gépet általában csak kevés külső gép támad egyszerre (a hálózatban a be-fokszám 10-nél alacsonyabb 4 kivételtől eltekintve), feltételezhetjük, hogy a külső gépek között nincs kapcsolat, pusztán véletlen egybeesés, hogy egy gépet a vizsgált időszakban többen is támadtak. Az utolsó hálózati tulajdonság, amire itt fel szeretnénk hívni a figyelmet, az a néhány magas fokszámú csúcs, amely egy IP-ről indul ki, és sok belső IP szkennelését jelzi. Itt már nem beszélhetünk véletlenszerű próbálkozásról, inkább a hálózat egy-egy fontosnak tartott szegmensének következetes feltárásáról és teszteléséről lehet szó. Érdekes, hogy ezek a nagy fokszámú, egyszerre sok gépet támadó IP-címek sem fednek át egymással a célpontjaikat illetően. Minden külső IP-címnek megvan a maga pásztázási tartománya, amelyek közt nincs átfedés.

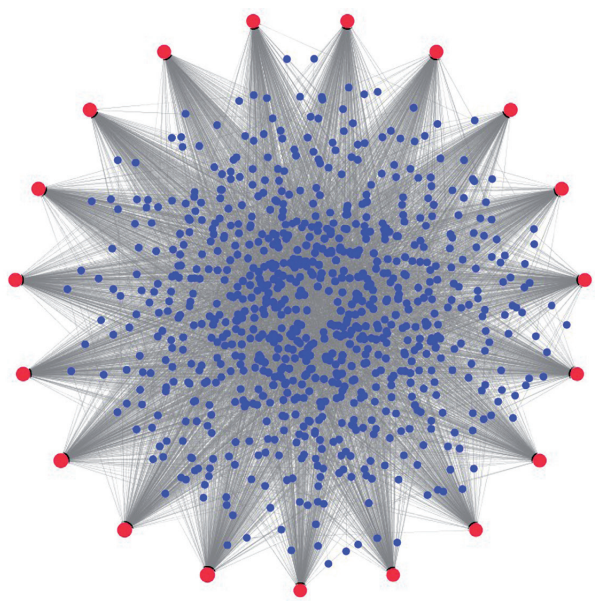
Megjegyezzük, hogy a hálózat töredezett jellege – azaz hogy nagyon sok külső címről, egymással átfedés nélkül viszonylag sok belső címet támadtak – felvetheti egy igen nagy skálájú támadás lehetőségét is, amellyel a teljes belső informatikai park felmérése és esetleges támadása történt. Bár ennek a globális támadásnak kicsi az esélye, hiszen a bemutatott példa minden egyes támadását egyedileg könnyű észlelni és így a támadást elhárítani. A következő alfejezetben viszont azt tárgyaljuk, hogyan lehet az összehangolt támadások egy típusát azonosítani.

Port-scan támadás IP-farmról

Az előző alfejezetben bemutattuk, hogy egyedi IP-címekről hogyan támadják a közszolgálati infrastruktúrát. Míg ott az egyetlen kiinduló IP-cím miatt a tűzfal könnyen azonosíthatta a támadót, a jelentősebb eszközparkkal rendelkező, rosszindulatú internetes szereplők úgynevezett IP-farmokat vetnek be a támadásra. Ezek az IP-farmok összehangoltan dolgoznak, hogy az egyedi próbálkozások kis kockázatú eseménynek látszódjanak, de az üzemeltető a teljes farm által begyűjtött adatok alapján mégis teljes képet alkothasson.

Egy IP-farm-támadás észleléséhez a következő feltételek együttes fennállását figyeltük:

- a forgalmi hálózatban a kapcsolat kezdeményezője külső IP-cím,
- egy hálózati élhez (külső IP → belső IP) csak kevés portszám fordul elő, azaz a külső gép csak kevés különböző portszámhoz csatlakozott a belső gépen,
- a külső IP-címekről érkező megkeresések célportszámait összegyűjtve egy belső IP-címen több mint 500 különböző portszámot próbáltak ki.



13. ábra

Port-scan-támadás IP-farmról

Forrás: a szerzők szerkesztése

Az IP-farm tagjai tehát felosztják egymás között, hogy melyik IP-ről melyik portokat fogják tesztelni. Az alábbi ábrán az ilyen tulajdonságot mutató IP-IP kapcsolatok hálózata látható. Jól észrevehető a 12-es ábrához képest a különbség. Itt egy-egy támadó IP egymás után több belső címmel is próbálkozik, és az egész hálózat egyetlen gyengén összekötött komponenst tartalmaz. A hálózat tipikusan magas fokszámú csúcsokból áll, és igen sűrűn összekötött. A forgalom időbeli lefolyásának és a pásztázott portszámok eloszlásának részletesebb elemzésével becslést adhatunk arra is, hogy vajon egyetlen IP-farm támadását látjuk-e, vagy több farm egyidejű, egymással átfedő adatokat gyűjtő támadásáról van-e szó.

Összefoglalás

A botnetek vonatkozásában a detektálási képességek, illetve az információbiztonság terén általános jelleggel a biztonsági kontrollok fejlődése a támadói apparátust követte és követi, ahogy e megállapítást az előző fejezetben tárgyaltuk. A botnetek felépítésének ismerete kritikus jelentőségű. *A védelmi képességek célzott fejlesztése* című alfejezetben tett megállapításoknak megfelelően a véletlen vagy skálafüggetlen modellel leírható felépítésű botnetek elleni védelmi képességek eltérő szükségletekkel, jellemzőkkel bírnak.

Ezért bár minden rendelkezésre álló lehetőséget és tudást (például mesterséges intelligencia, fuzzy logika) be szükséges vetni a botnetek *a priori* detektálásába, és a jelenlegi tudáshalmazra épülő algoritmusok fejlesztése is szükségszerű, a botnetek *a posteriori* detektálása jelenti a tényleges megoldást. Ez utóbbi esetben az adott botnet működési jellemzőinek (alkalmazott architektúra, hálózati protokollok, képességek) ismerete segítséget jelent a detektálásban.

A Mirai botnet vonatkozásában a hálózat 2016 augusztusában lépett működésbe, szeptemberben pedig már DDoS-támadás következtében a címlapokon szerepelt. A megalkotóját mindössze 2017 elején sikerült azonosítani, továbbá a 2016. novemberben kivitelezett Deutsche Telekom elleni támadásért felelős személyt 2017. februárban tartóztatták le.[47] Azonban a botnet analizálása ezt követően is zajlott, amely kiválóan szemlélteti a műveletek időbeli szükségleteit és az emiatt felmerülő hiányosságokat. Ebből következik, hogy a botnet megjelenését követő detektálás és válaszreakció kivitelezése között eltelt időintervallum kritikus jelentőségű, minimalizálása megkerülhetetlen.

Felhasznált irodalom

Acaralia, D. – Rajarajana, M. – Komninos, N. – Herwono, I. (2016): Survey of approaches and features for the identification of HTTP-based botnet traffic. *Journal of Network and Computer Applications,* Vol. 76. 1–15.

Akamai (2015): *XOR DDoS Botnet Launching 20 Attacks a Day From Compromised Linux Machines, Says Akamai.* Elérhető: www.akamai.com/us/en/about/news/press/2015-press/xor-ddos-botnet-attacking-linux-machines.jsp (A letöltés dátuma: 2019. 03. 21.)

[47] Antonakakis et al. 2017, 2.

ALPCAN, T. – BAŞAR, T. (2011): *Network Security. A Decision and Game-Theoretic Approach.* Cambridge, Cambridge University Press.

ANTONAKAKIS, M. – APRIL, T. – BAILEY, M. – BERNHARD, M. – BURSZTEIN, E. – COCHRAN, J. – DURUMERIC, Z. – HALDERMAN, J. A. – INVERNIZZI, L. – KALLITSIS, M. – KUMAR, D. – LEVER, C. – MA, Z. – MASON, J. – MENSCHER, D. – SEAMAN, C. – SULLIVAN, N. – THOMAS, K. – ZHOU, Y. (2017): *Understanding the Mirai Botnet.* Elérhető: http://mdbailey.ece.illinois.edu/publications/usesec17_mirai.pdf (A letöltés dátuma: 2019. 03. 21.)

BARABÁSI A.-L. – BONABEA, E. (2003): Scale-free Networks. *Scientific American,* Vol. 288, No. 5. 50–59. Elérhető: http://barabasi.com/f/124.pdf (A letöltés dátuma: 2019. 03. 21.)

BARABÁSI A.-L. (2001): The Physics of the Web. *Physics World,* Vol. 14, No. 7. 33–38.

BEDERNA Zs. (2013): Ködös biztonság. *IT Business,* Vol. 11, No. 11. 22–23.

BEDERNA Zs. (2015): Fuzzy-based intrusion detection. *Hadmérnök,* Vol. 10, No. 1. 147–160.

BIJALWAN, A. – CHAND, N. – PILLI, E. S. – RAMA KRISHNA, C. (2016): Botnet analysis using ensemble classifier. *Perspectives in Science,* Vol. 8. 502–504. DOI: https://doi.org/10.1016/j.pisc.2016.05.008

BOSHMAF, Y. – MUSLUKHOV, I. – BEZNOSOV, K. – RIPEANU, M. (2013): Design and analysis of a social botnet. *Computer Networks,* Vol. 57, No. 2. 556–578. DOI: https://doi.org/10.1016/j.comnet.2012.06.006

CANTÓN, D. (2015): Botnet detection through DNS-based approaches. *INCIBE,* 2015. 01. 20. Elérhető: www.incibe-cert.es/en/blog/botnet-detection-dns (A letöltés dátuma: 2019. 03. 21.)

CHEN, C.-M. – LIN, H.-C. (2015): Detecting botnet by anomalous traffic. *Journal of Information Security and Applications,* Vol. 21. 42–51. DOI: https://doi.org/10.1016/j.jisa.2014.05.002

CHEN, W. – LUO, X. – YIN, C. – XIAO, B. – HO AU, M. – TANG, Y. (2017): CloudBot. Advanced mobile botnets using ubiquitous cloud technologies. *Pervasive and Mobile Computing,* Vol. 41. 270–285. DOI: https://doi.org/10.1016/j.pmcj.2017.03.007

CORREIA, P. – ROCHA, E. – NOGUEIRA, A. – SALVADOR, P. (2012): Statistical Characterization of the Botnets C&C Traffic. *Procedia Technology,* Vol. 1. 158–166. DOI: https://doi.org/10.1016/j.protcy.2012.02.030

Cybercrime as a Service (2013). *General Security,* 08. 07. Elérhető: https://resources.infosecinstitute.com/cybercrime-as-a-service/ (A letöltés dátuma: 2019. 03. 21.)

DAGON, D. – GU, G. – LEE, C. P. – LEE, W. (2007): A Taxonomy of Botnet Structures. In *Twenty-Third Annual Computer Security Applications Conference (ACSAC 2007).* Miami Beach, IEEE. 325–339. DOI: https://doi.org/10.1109/ACSAC.2007.44

ENISA: *Botnets* (é. n.). Elérhető: www.enisa.europa.eu/topics/csirts-in-europe/glossary/botnets (A letöltés dátuma: 2019. 03. 21.)

FEILY, M. – SHAHRESTANI, A. – RAMADASS, S. (2009): A Survey of Botnet and Botnet Detection. In *Third International Conference on Emerging Security Information, Systems and Technologies.* Athens–Glyfada, IEEE. 268–273. DOI: https://doi.org/10.1109/SECURWARE.2009.48

FERGUSON, R. (2010): *The Botnet Chronicles: A Journey to Infamy.* Cupertino, Trend Micro Inc. Elérhető: www.trendmicro.co.kr/cloud-content/us/pdfs/security-intelligence/white-papers/wp_botnet-chronicles.pdf (A letöltés dátuma: 2019. 03. 21.)

GARCÍA, S. – GRILL, M. – STIBOREK, J. – ZUNINO, A. (2014): An empirical comparison of botnet detection methods. *Computers & Security,* Vol. 45. 100–123. DOI: https://doi.org/10.1016/j.cose.2014.05.011

Hyslip, T. – Pittman, J. (2015): A Survey of Botnet Detection Techniques by Command and Control Infrastructure. *Journal of Digital Forensics, Security and Law,* Vol. 10, No. 2. 7–26. DOI: https://doi.org/10.15394/jdfsl.2015.1195

IBM Corporation (2016): *The inside story on botnets.* Elérhető: www.ibm.com/downloads/cas/V3YJVYZX (A letöltés dátuma: 2019. 03. 21.)

IBM Corporation (2017): *The Weaponization of IoT Devices.* Elérhető: www.ibm.com/downloads/cas/6MLEALKV (A letöltés dátuma: 2019. 03. 21.)

Karim, A. – Salleh, R. – Shiraz, M. – Shah, S. – Awan, I. – Anuar, N. (2014): Botnet detection techniques. Review, future trends, and issues. *Journal of Zhejiang University – Science C,* Vol. 15, No. 11. 943–983. DOI: https://doi.org/10.1631/jzus.C1300242

Khattak, S. – Ahmed, Z. – Syed, A. A. – Khayam, A. (2015): BotFlex. A community-driven tool for botnet detection. *Journal of Network and Computer Applications,* Vol. 58. 144–154. DOI: https://doi.org/10.1016/j.jnca.2015.10.002

Kirubavathi, G. – Anitha, R. (2016): Botnet detection via mining of traffic flow characteristics. *Computers and Electrical Engineering,* Vol. 50. 91–101. DOI: https://doi.org/10.1016/j.compeleceng.2016.01.012

Mathur, L. – Raheja, M. – Ahlawat, P. (2018): Botnet Detection via mining of network traffic flow. *Procedia Computer Science,* Vol. 132. 1668–1677. DOI: https://doi.org/10.1016/j.procs.2018.05.137

McAfee (2013): *Cybercrime Exposed. Cybercrime as a Service.* Elérhető: www.unife.it/ing/lm.infoauto/sicurezza-si/files/materiale-extra-2017-2018/cybercrime-exposed (A letöltés dátuma: 2019. 03. 21.)

Rrushi, J. – Mokhtari, E. – Ghorbani, A. A. (2011): Estimating botnet virulence within mathematical models of botnet propagation dynamics. *Computers & Security,* Vol. 30, No. 8. 791–802. DOI: https://doi.org/10.1016/j.cose.2011.07.004

Goncharov, M. (2012): *Russian Underground 101.* Research Paper. Cupertino, Trend Micro. Elérhető: www.trendmicro.de/cloud-content/us/pdfs/security-intelligence/white-papers/wp-russian-underground-101.pdf (A letöltés dátuma: 2019. 03. 21.)

Silva, S. S. C. – Silva, R. M. P. – Pinto, R. – Salles, R. M. (2013): Botnets. A Survey. *Computer Networks,* Vol. 57, No. 2. 378–403. DOI: https://doi.org/10.1016/j.comnet.2012.07.021

Siciliano, R. (2011): 7 Types of Hacker Motivations. *McAfee,* 2011. 03. 16. Elérhető: http://blogs.mcafee.com/consumer/identity-theft/7-types-of-hacker-motivations (A letöltés dátuma: 2019. 03. 21.)

Szőr P. (2010): *A vírusvédelem művészete.* Bicske, SZAK Kiadó.

Kudo, T. – Kimura, T. – Inoue, Y. – Aman, H. – Hirata, K. (2018): Stochastic modeling of self-evolving botnets with vulnerability discovery. *Computer Communications,* Vol. 124. 101–110. DOI: https://doi.org/10.1016/j.comcom.2018.04.010

Wang, P. – Sparks, S. – Zou, C. C. (2010): An Advanced Hybrid Peer-to-Peer Botnet. *IEEE Transactions on Dependable and Secure Computing,* Vol. 7, No. 2. 113–127. DOI: https://doi.org/10.1109/TDSC.2008.35

Zhao, D. – Traore, I. – Sayed, B. – Lu, W. – Saad, S. – Ghorbani, A. – Garant, D. (2013): Botnet detection based on traffic behavior analysis and flow intervals. *Computers & Security,* Vol. 39. Part A. 2–16. DOI: https://doi.org/10.1016/j.cose.2013.04.007

Ludovika Egyetemi Kiadó Nonprofit Kft.
Székhely: 1089 Budapest, Orczy út 1.
Kapcsolat: info@ludovika.hu

A kiadásért felel: Koltányi Gergely ügyvezető igazgató
Felelős szerkesztő: Inzsöl Kata
Olvasószerkesztő: Szarvas Melinda
Korrektor: Szabó Ilse
Tördelőszerkesztő: Stubnya Tibor

Nyomdai kivitelezés: Pátria Nyomda Zrt.
Felelős vezető: Orgován Katalin vezérigazgató

ISBN 978-963-531-080-7 (nyomtatott)
ISBN 978-963-531-081-4 (elektronikus)

A hálózatkutatás korunk egyik kurrens kutatási témája, azonban alkalmazása számos területen még csak ötlet szintjén található meg. A hálózatkutatás alapvető eleme az adat, amelyen vagy amellyel kutatásokat lehet folytatni, de önmagában nem elégséges az eredmények eléréséhez, amihez hálózattudományi ismeretek is szükségesek. Ugyanakkor még talán ezen kettős elemi feltétel megléte esetén sem várhatunk eredményeket. Létfontosságú az a szakmai háttér is, amely megfogalmazza a hipotéziseket, segít eligazodni az adatok által kirajzolt mintákon, eredményeken, és végső soron hasznosítani tudja azok tanulságait.

A kötet a közigazgatás szerteágazó területeiről mutat be esettanulmányokat, konkrét problémák megoldására nyújt modelleket, valamint szemlélteti a módszer alkalmazási lehetőségeit. A kutatócsoport reményei szerint a tanulmányok egy jövőbeni alkalmazott hálózatkutatási tantárgy oktatási segédanyagaként is szolgálhatnak majd.

SZÉCHENYI 2020

MAGYARORSZÁG KORMÁNYA

Európai Unió
Európai Szociális Alap

BEFEKTETÉS A JÖVŐBE