

# A TÁRSADALOM HÁLÓZATI JELENSÉGEINEK ADATVEZÉRELT VIZSGÁLATA

## DATA DRIVEN INVESTIGATION OF NETWORK PHENOMENA OF SOCIETY

Kertész János<sup>1</sup>, Roberta Sinatra<sup>2</sup>, Vedres Balázs<sup>3</sup>

<sup>1</sup>az MTA rendes tagja, Kerteszl@ceu.edu

<sup>2</sup>PhD, SinatraR@ceu.edu,

<sup>3</sup>PhD, VedresB@ceu.edu

Közép-európai Egyetem (CEU) Hálózattudományi Központ

### ÖSSZEFOGLALÁS

A CEU Hálózattudományi Központja társadalmilag releváns kérdések adatokon alapuló, hálózati szempontú vizsgálatával foglalkozik. Néhány példa: Milyen az emberi kapcsolatok szerkezete társadalmi méretekben a kommunikációs adatok tükrében, és hogyan változik mindez az új médiumok hatására? Hogyan terjednek a járványok, az információ vagy egy innováció a társadalomban? Hogyan lehet optimalizálni egy team összetételét a kreativitás szempontjából? Mi a nemek szerepe a munkacsoportok hatékonyságában? Hogyan lehet mérni és előre jelezni a sikert? Ebben a cikkben röviden áttekintünk néhány eredményt, amelyeket a társadalmi terjedési jelenségek, a teamek kreativitásának hálózati vonatkozásai és a siker társadalmi elmélete területeit kutatva értünk el.

### ABSTRACT

The Center for Network Science of the Central European University (CEU) deals with data-based, network related investigation of socially relevant problems. Some examples: What is the structure of human interactions at the societal scale in the light of communication data and how does it change due to the new media? How do epidemics, information or an innovation spread in society? How can the composition of a team be optimized from creativity's point of view? What is the role of genders in the efficiency of working teams? How can success be quantified and predicted? In this article we briefly survey some results we have obtained by researching social contagion phenomena, the network aspects of team creativity and the social theory of success.

**Kulcsszavak:** hálózattudomány, adatvezérelt kutatás, társadalmi terjedési jelenségek, munkacsoportok kreativitása, a siker mérése és modellezése

**Keywords:** network science, data driven research, social contagion, creativity of working teams, measuring and modelling success

## ADATÖZÖN, KOMPLEXITÁS ÉS HÁLÓZATI SZEMPONT A TÁRSADALOM VIZSGÁLATÁBAN

A társadalom a komplex rendszerek paradigmatis példája: nagyszámú kölcsönható egyedből épül fel, visszacsatolások, nem-linearitások jellemzik, és úgynevezett emergens jelenségek lépnek fel benne. Az ilyen rendszerekre érvényes Philip W. Anderson Nobel-díjas fizikus híres mondása: „A több másmilyen” („More is different”), amivel a természettudományokat évszázadokon keresztül meghatározó redukcionizmus korlátaira mutatott rá. Nem elég alaposan megérteni az alkotóelemek – a társadalom esetében: az emberek – közötti kölcsönhatásokat (és persze az alapos megértéstől távol vagyunk), az alkotóelemek nagy száma új minőséget hozhat létre. Ilyenek lehetnek a nagy társadalmi változások, az új struktúrák, intézmények megjelenése, de akár egy innováció létrejötte és elterjedése is. Stephen Hawking szerint a 21. század a tudományban a komplexitás évszázada lesz. Ennek a jóslatnak a beteljesüléséhez új módszerekre, új megközelítésekre és diszciplínákon átívelő együttműködésre van szükség.

A társadalom az egyik legkomplexebb rendszer, mégis, az új tudományos erőfeszítések jelentős része rá irányul. Ennek oka részben abban rejlik, hogy a globális világunkat érő kihívásoknak úgy tudunk megfelelni, ha jobban megértjük a társadalom szerkezetét és dinamikáját. Mindehhez az ezredfordulóra értek meg azok a technikai feltételek, amelyek a tudomány gyors fejlődését lehetővé tették.

Az elmúlt két évtized drámai változást hozott a társadalmi jelenségek kvantitatív vizsgálatában. Az infokommunikációs technológia fejlődése, a digitális megoldások általánossá válása a legkülönbözőbb területeken hihetetlen mennyiségű adatot eredményez. Ma már az emberi aktivitás önkéntelenül is folyton adatokat hoz létre az élet szinte minden területén, legyen szó kommunikációról, egészségügyi vizsgálatokról, kereskedelmi tevékenységről vagy közlekedésről. Ezek a „digitális lábnyomok” alkalmasak az egyéni viselkedés részletekbe menő, „mikroszkopikus” tanulmányozására, ráadásul óriási, társadalmi méretű („makroszkopikus”) mintákon. Olyan ez, mintha egy gáz viselkedését nem a hőmérséklete, sűrűsége és nyomása segítségével, hanem az egyes molekulákat nyomon követve próbálnánk leírni. A gázok esetében tudjuk, hogyan jönnek létre a makroszkopikus állapotjelzők a mikroszkopikus kölcsönhatásokból és a mozgástörvényekből, így a molekuláris szintű leírás gyakran fölösleges. A társadalom esetében a helyzet bonyolultabb: a mikroszkopikus mozgástörvényeket és a kölcsönhatásokat nem vagy alig ismerjük, így felbecsülhetetlen értékű információkhoz juthatunk az adatözön által biztosított, új lehetőségek révén.

Az ezredforduló körüli forradalmi változásokat a társadalom, és általában a komplex rendszerek vizsgálatában nemcsak az adatözön (Big Data), hanem egy attól nem független szemléleti megújulás is okozta. Ennek elindítója két elméleti dolgozat volt: Duncan J. Watts és Steven H. Strogatz (1998) kisvilág-hálózatokról, valamint Barabási Albert-László és Albert Réka (1999) skálamentes hálózati

modellről szóló cikke. A figyelem a komplex rendszerek váza, a mögöttük rejlő hálózatok felé fordult, és kiderült, hogy tanulmányozásuk elengedhetetlen a komplex rendszerek megértéséhez. Az egyik legmeglepőbb felismerés, hogy a komplex rendszerek sokfélesége ellenére hálózataik számos univerzális sajátosságot mutatnak. Ilyen jellemzők a hálózatok méretéhez képest kis átlagos távolság (lépésszám), a csomópontok erősen heterogén kapcsolódásai (fokszám) és a moduláris szerkezet. Mára a hálózattudomány önálló diszciplínává vált, és kiváló tankönyvek tárgyalják a komplex rendszerek hálózatainak tulajdonságait, a rajtuk zajló folyamatokat és modellezésüket.

A Közép-európai Egyetemen (CEU) 2008-ban alakult Hálózattudományi Központ célul tűzte ki a társadalmilag releváns kérdések adatokon alapuló, hálózati szempontú vizsgálatát. A fentiekből is érzékelhető, hogy a kutatómunka első számú feltétele a megfelelő adatokhoz való hozzáférés, vagy megfordítva: azok a kutatások válnak lehetségessé, amelyekhez sikerül adatokat szerezni. Ilyen, rendelkezésünkre álló adatok például a mobiltelefon-forgalomra vonatkozó számlázási adatok, a Skype internetes rendszer hálózati felépülési adatai, az iWiW internetes szociális hálózat adatai, a *Web of Science* adatai, az amerikai dzsesszenészek együtteseire és lemezfelvételeire, valamint a számítógépes játékok fejlesztő *team*-jeire és a termékek sikerességére vonatkozó adatok. Az ilyen mélységű adattömeg aranybánya a tudomány számára, mindazonáltal használata különleges körülményt igényel. A projekteket ezért a kutatóhelyek etikai bizottságai felügyelik, és biztosítják, hogy a személyiségi jogok ne kerüljenek konfliktusba a kutatással, és így van ez a CEU esetében is.

A továbbiakban néhány példát mutatunk be a CEU adatvezérelt kutatásaiból. Először a társadalmi terjedési jelenségekkel foglalkozunk innovációk terjedésének példáján. Utána néhány eredményt mutatunk be azzal kapcsolatban, hogy a csapatok összetétele hogyan befolyásolja a kreativitásukat. Végül a társadalmi sikeresség új elméletét tárgyaljuk röviden a tudományos eredményesség ismert mérőszámaiból kiindulva. A cikket egy rövid kitekintéssel zárjuk.

## TÁRSADALMI TERJEDÉSI JELENSÉGEK

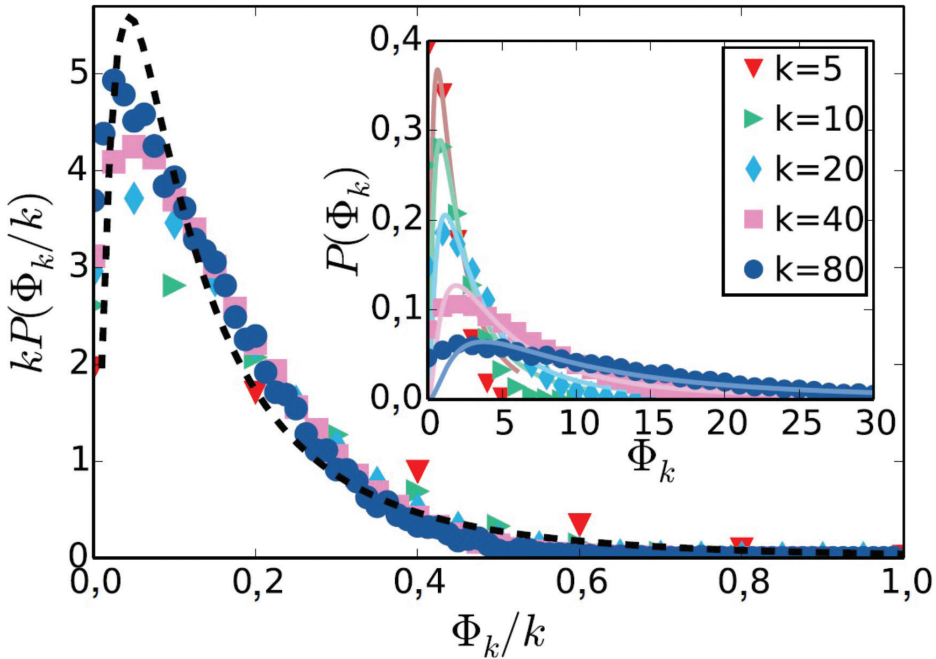
Az emberi kapcsolatok hálózatán számos terjedési jelenséget figyelhetünk meg. A hírek, pletykák, viselkedési formák vagy innovációk jelentős mértékben ilyen módon érik el a társadalom egyre nagyobb hányadát, de a járványok terjedése is ide sorolható. Az információ és a járványok terjedése közötti analógiára már a hatvanas években felhívták a figyelmet. Az adattudomány és a hálózatok kutatás együttes erőfeszítéseinek köszönhetően a járványterjedés leírásában, sőt az előrejelzésében is jelentős haladást sikerült elérni az elmúlt években (Chovell et al., 2017). Az említett társadalmi jelenségek terjedésének leírása azonban bizonyos

értelemben még nehezebb. Itt ugyanis nem valamilyen anyagi hordozók (mikrobák) párkölcsönhatásokban megnyilvánuló átadásáról van szó, hanem döntésekről és azokat megelőző hatásokról. Elhiszem-e és továbbadom-e a pletykát? Átveszem-e az adott viselkedési formát? Megveszem-e az új technikai berendezést? Ezek a döntések általában a környezet, az ismerősök állapotától, hozzáállásától függenek. Az innováció példájánál maradva: ha ismerőseim jelentős része egy új típusú telefonkészüléket használ, és jó véleménnyel van róla, én is hajlok rá, hogy beszerezzek egyet.

Az ilyen komplex terjedési folyamatok (complex contagion processes) leírására vezette be Mark S. Granovetter a küszöbmodellt, amit Duncan J. Watts (2002) öntött matematikailag jól kezelhető formába. A modell lényege a következő: a személyek egy hálózat csúcspontjai, akik társadalmi kapcsolataikkal vannak összekötve. Minden  $i$  egyed rendelkezik egy  $\phi_i$  küszöbértékkel, ami az innovációt adaptáló szomszédok kritikus hányadát jelzi. Vagyis, ha  $k_i^{(a)}/k_i \geq \phi_i$ , akkor az  $i$ -edik egyed is adaptálni fogja az innovációt, ahol  $k_i^{(a)}$  az adaptáló,  $k_i$  pedig az összes szomszédok száma, a fokszám. Az innovációt egy véletlenül kiválasztott csúcsponttól indítjuk. (Itt az innováció helyettesíthető más, társadalmilag terjedő jelenséggel.) Ha egy csúcspont körül az adaptáló szomszédok aránya túllépi a küszöbértéket, akkor a csúcspont maga is adaptálónak válik, ezáltal egy addig nem adaptáló szomszédjánál következhet be a küszöb túllépése, és így tovább. Ilyen módon adaptációs kaszkádok, vagyis egymást előidéző követők füzerei jöhetnek létre.

A modellben a hálózat jellege, az átlagos küszöbérték és a fokszám függvényében két fázist lehet megkülönböztetni, egy globális és egy lokális terjedési fázist. A paraméterek egy tartományában globális kaszkádok alakulhatnak ki – ez felel meg a gyors terjedésnek. A másikban a kaszkádok elhalnak, globális terjedés nem lehetséges. Ez az egyszerű modell számot ad arról az általános megfigyelésről, hogy bizonyos innovációk futótűzként elterjednek a társadalomban, míg mások kudarcot vallanak.

A Watts-féle küszöbmodell azonban nem alkalmas a megfigyelési adatok közvetlen értelmezésére. Igaz, a társadalmi terjedési jelenségekről csak a legutóbbi időben sikerült részletes adatokhoz hozzáférni. Kutatócsoportunk a Skype „Voice over Internet” rendszer adatait vizsgálta (Karsai et al., 2016). Rendelkezésünkre álltak az anonimizált felhasználók regisztrációs adatai, a visszajelzett kapcsolatok kiépítésének és az utolsó belépésnek az időpontja – az ingyenes és a „fizetős” szolgáltatási platformokon egyaránt. Ez utóbbi jelentősége, hogy a fizetős szolgáltatások mindig az ingyenes hálózaton terjednek, így a háttérhálózat ebben az esetben ismert – ellentétben azzal a helyzettel, amikor az innováció az egyébként ismeretlen társadalmi kapcsolatok hálózatán terjed. Ilyen fizetős szolgáltatások tanulmányozása révén lehetett a küszöbérték-eloszlásról információt szerezni (1. ábra).



**1. ábra.** Betét: A szolgáltatást már használó szomszédok  $\Phi_k$  számának eloszlása a szolgáltatás adaptációjakor különböző fokszámok esetében, amelyek eltérő szimbólumokkal vannak jelölve.

Nagy ábra: Ugyanaz, de a szolgáltatást már használó szomszédok arányának ( $\Phi_k/k$ ) függvényében. A folytonos, illetve szaggatott vonalak illetett lognormál eloszlásokat mutatnak (Karsai et al., 2016 nyomán)

Azzal, hogy a nagy ábrán a görbék jó közelítéssel egymásra esnek, sikerült bebizonyítani az elmélet korábbi vitatott feltételezését, hogy a küszöbérték megfelelő változója nem az adaptáló szomszédok száma, hanem aránya (Karsai et al., 2016).

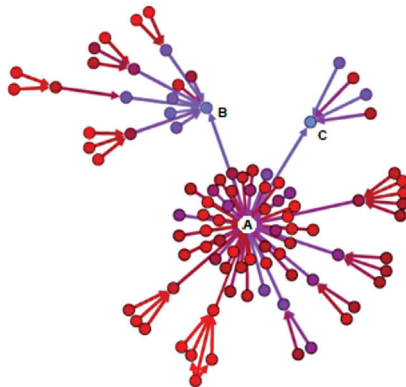
Az adatok arra is alkalmasak, hogy a terjedés mechanizmusába mélyebben betekintsünk. Megállapítható volt, hogy – ellentétben a Watts-modell által sugalmazott képpel – úgy is létrejöhet globális terjedés, hogy nem alakul ki globális kaszkád. Nyilvánvaló, hogy a Watts-modellből hiányoznak bizonyos fontos komponensek. Az egyik kézenfekvő: a terjedést nemcsak a szomszédok viselkedése, illetve a tőlük érkező információ befolyásolja, hanem külső hatások is, például hírek a médiában vagy reklámkampányok. Ilyen hatások révén „spontán” adaptálók léphetnek fel, akik aztán új kaszkádok kiindulópontjai lehetnek. Ez a mechanizmus globális terjedést eredményez kis kaszkádok sorozatán keresztül, ahogyan az az adatokon is megfigyelhető.

Egy másik, az eredeti küszöbmodellekből hiányzó vonás az innovációtól teljesen elzárkózók figyelembevétele. Vannak, akik sem az ismerőseik, vagyis há-

lózati szomszédjaik példáját követve, sem pedig külső hatásra nem hajlandók az innovációt adaptálni. Ennek számos oka lehet, kezdve azon, hogy elégedettek egy korábbi technikai megoldással, addig hogy ideológiai (például környezetvédelmi) szempontok miatt nem hajlandók kipróbálni az új terméket. Az ilyenek beleszámítanak az adaptáló szomszédok arányába, és lassítják a terjedési folyamatot.

A fenti két jelenséget sikerült egy általánosított küszöbmodellbe beépíteni, és a modellt bizonyos típusú hálózatokon jó közelítő módszerekkel megoldani (Ruan et al., 2015). A modell paramétereit a megfigyelésekből lehetett származtatni, kivéve az elzárkózók arányát, amire vonatkozó adatunk természetesen nem áll rendelkezésre, úgyhogy ezt a paramétert a modell kalibrálása során illesztettük. A modell igazolását a bemenő adatoktól nem közvetlenül függő mennyiségeknek a megfigyeltekkel való összehasonlítása jelenti, így például a kaszkádok eloszlása (Karsai et al., 2016).

A fent bemutatott mechanizmus nemcsak innovációk terjedését írja le, hanem alkalmazható például az *online* szociális hálózatok összeomlásának modellezésére is (Török–Kertész, 2017). Ismeretes, hogy a korábban Magyarországon rendkívül népszerű iWiW-szolgáltatás nagyon rövid idő alatt lényegében összeomlott. A Facebooknál korábban, 2002-ben alapított iWiW-et végül 2014 júniusában végképp leállították. Egy online szociális hálózat elveszti vonzerejét, ha az ismerősök jelentős hányada már máshol aktív – ez az a kollektív hatás, amelyet a küszöbmodellek vesznek figyelembe. Ebben a rendszerben az erős vetélytársról, a Facebookról érkező hírek jelentik a külső hatást. Annak ellenére, hogy itt a „makacs ragaszkodók” aránya jóval kisebb volt az előző példában bemutatott  $r$  értéknél, itt is véges kaszkádok alakultak ki, amelyeket a spontán távozók váltottak ki. A 2. ábra egy ilyen véges kaszkádot mutat be.



**2. ábra.** Az iWiW online szociális hálózatról távozók egy véges kaszkádja. A nyilak a befolyást jelzik, vagyis a korábban távozóról a később távozóra mutatnak. Először A egy kaszkád eredményeképpen távozik, hozzájárulva B és C későbbi távozásához.

(Török–Kertész, 2017 nyomán)

## MUNKACSOPORTOK KREATIVITÁSÁNAK HÁLÓZATI VONATKOZÁSAI

Az innováció terjedését meg kell hogy előzze annak létrehozása, ami egy kreatív folyamat. Ez legtöbbször munkacsoportokban, teamekben zajlik, így kézenfekvő a kérdés: hogyan befolyásolja a teamek kreativitását az összetételük és a tagjaik közötti kapcsolatrendszer? Az innováció terjedésének irodalmához képest az innováció létrehozásának hálózati feltételeiről meglepően kevés munka szól. A kérdés hálózati irodalmában az uralkodó paradigma szerint a hálózatok szerepe kettős: az újításokhoz egyrészt szükség van gyenge, de távoli hálózati régiókba mutató kötésekre, és erős, csoporton belüli kötésekre. E paradigma szerint a gyenge és diverz kötések jelentősége abban áll, hogy ötleteket csatornázzanak be kívülről egy munkacsoportba, míg az erős és homogén kötések az ötletek megvalósításában játszanak fontos szerepet (Aral–Alstyn, 2011; Obstfeld, 2005).

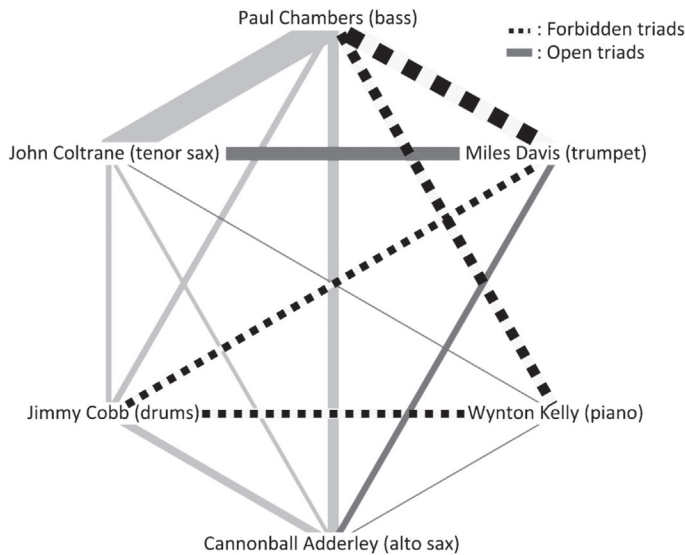
A kreativitás gyenge és erős kötésekre építő irodalma azonban éppen a kreativitás alapkérdését hagyja megválaszolatlanul: hogyan járulnak hozzá kapcsolataink az új ötletek keletkezéséhez? A gyenge kötések, amelyek sokféle társadalmi miliőhöz kapcsolhatnak, valóban alkalmasak arra, hogy eladdig ismeretlen ötletekhez adjanak hozzáférést, de mi magyarázza ezeknek az ötleteknek a keletkezését?

Korábbi munkáinkban erre a kérdésre válaszul alkottuk meg a strukturális gyűrődések fogalmát (De Vaan et al., 2015; Vedres–Stark, 2010). A strukturális gyűrődések egy kontextuson (például munkacsoporton) belüli átfedések kohézív csoportok között. Ezek az átfedések hozzájárulnak ahhoz, hogy a munkacsoport tagjai egyrészt felfedezzenek lehetséges új kombinációkat, másrészt (a már meglévő kohézióknak köszönhetően) meg is tudják valósítani az új kombinációkat mint új terméket, kutatási tervet, műalkotást.

A strukturális gyűrődések hipotézise átfedő csoportok esetében egyedibb termékeket feltételez, illetve sikeresebb termékeket (amennyiben a mező kulcsértékei közé tartozik a kreativitás). Átfedő csoportokat azonban nem könnyű azonosítani: ehhez egyrészt szükség van egy megfelelően nagy csoportra, ahol „van hely” két alcsoport számára. (Ha az elemi kohézív csoport három főből áll, akkor legalább ötfős csoportokra van szükség, hogy két háromfős alcsoportot azonosítsunk, amelyek egy csomóponton átfednek.) Másrészt arra van szükség, hogy a munkacsoport belső hálózata ne legyen túl sűrű, mert ebben az esetben két alcsoport nem lesz azonosítható.

Ezen kihívásokat orvosolandó, javasoltuk a tiltott triádok sűrűségét mint a strukturális gyűrődések elemi mutatóját. A tiltott triádok olyan nyitott triádok, ahol a két meglévő él súlya nagy (Granovetter, 1973). Két átfedő kohézív csoport átfedése körül számos tiltott triád található: az élék súlya nagy (azaz a kötések erősek), mivel minden él kohézív kontextusban szerepel, ugyanakkor számos hiányzó él is található (azon csomópontok között, amelyek különböző kohézív csoportba tartoznak).

A dzsesszvilág egy olyan mező, ahol a kreativitás kifejezetten központi érték. Ezt a világot vizsgáltuk meg a *Tom Lord Discography* adatai alapján (Lord, 2010), amely a dzsessz legteljesebb diszkografikus adatbázisának mondható. Ezen adatok alapján a dzsessz teljes történetét elemeztük, 1896 és 2010 között, 175 064 lemezfelvételt, 42 929 zenekartól, összesen 187 784 zenész részvételével. Minden lemezfelvétel esetében rekonstruáltuk az abban szereplő zenészek korábbi együttműködési hálózatát. A 3. ábra Miles Davis és zenésztársainak súlyozott együttműködési hálózatát mutatja a *Kind of Blue* első *session*-je idején.



3. ábra. Miles Davis és zenésztársainak együttműködési hálózata a *Kind of Blue* felvételek (1959. március 2.) (Vedres, 2017 nyomán)

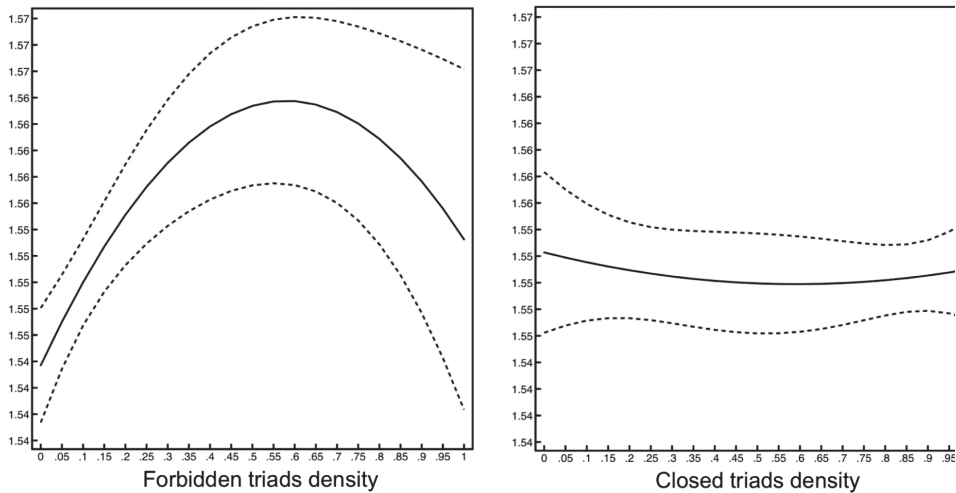
A 3. ábrán látható hálózat meglehetősen sűrű, mégis szerepel benne egy hiányzó él: Miles Davis és Wynton Kelly korábban soha nem játszottak együtt stúdióban, holott mindkettőjüknek vannak közös és erős kötésekkkel kapcsolatos zenésztársaik. Az ábrán szaggatott vonallal jelölt triádok Davis és Kelly körül tiltott triádok.

A sikeresség és a tiltott triádok sűrűsége között pozitív kapcsolatot feltételezhetünk a dzsessz esetében, bár a túlzott sűrűség már feltétlenül jelent sikert. Túlzottan sok tiltott triád esetén már a csoport koordinációja válik nehézkessé.

A várt összefüggést statisztikai modellekkel vizsgáltuk. A sikerességet a lemezfelvétel anyagából megjelenő kiadások számával mértük. Miles Davis *Kind of Blue* felvételéből például 186 kiadás született a későbbiek során (és ezzel ez a legsikeresebb dzsessz lemezfelvétel). Mivel a kiadások száma diszkrét eseményadatnak tekinthető, és a megoszlása szignifikánsan ferdebb, mint a Pois-



son-megoszlás, negatív binomiális regressziós modellt használtunk a becsléshez. A 4. ábra mutatja az eredményeket. A részletes modellek Vedres Balázs cikkében (2017) található. Ezen az ábrán a tiltott triádok és a zárt triádok marginális hatásait tüntettük fel, mialatt az összes kontrollváltozó az átlagánál volt rögzítve.



4. ábra. A tiltott és zárt triádok sűrűsége és a siker közötti összefüggés negatív binomiális regressziós modell alapján (Vedres, 2017 nyomán)

Az eredmények arra mutatnak, hogy a tiltott triádok hozzájárulnak a sikerességhez a dzsesszben, míg a zárt triádok nem. Azok a felvételek, ahol egy zenész több összeszokott zenésztársát is szerepelteti, akik még nem játszottak együtt, sikerebbek lehetnek, mint azok a zenészek, akik ugyanabban a felállásban többször is játszottak már együtt.

#### A SIKER MÉRÉSE ÉS MODELLEZÉSE

Mitől függ, hogy egy tudományos dolgozat, egy művészeti alkotás vagy egy innováció sikeres lesz-e? A *teljesítmény* az egy területen kifejtett aktivitás teljes, objektív eredményeinek összessége, ami például egy tudós publikációiban vagy egy sportoló, illetve egy csapat elért csúcsaiban, helyezéseiben jut kifejezésre, és egyéni tevékenységeken alapul. Ezzel szemben a hírnév, a népszerűség, a befolyás és a láthatóság által jellemzett *siker* kollektív jelenség, ami a közösségnek az egyén vagy egy csoport teljesítményére adott reakcióját, illetve annak elfogadását testesíti meg (Uzzi, 2008; Yucesoy–Barabási, 2016; Ke et al., 2015). Ezen két jelenség között a kapcsolatot gyakran nyilvánvalónak tekintik, pedig azt távolról

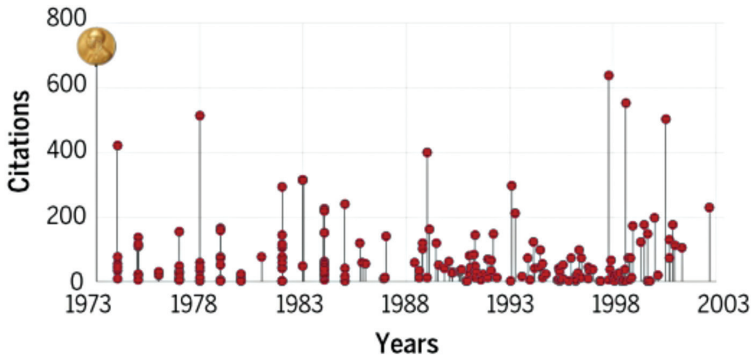
sem értjük, és gyakran ellentmondásokba ütközünk. Valóban, még a legalapvetőbb tudományos felfedezés is észrevétlen marad, ha a tudományos közösség nem ismeri el a fontosságát diszkussziók és hivatkozások formájában. A siker szociológiájának kiinduló feltételezése a teljesítmény és a siker közötti különbség felismerésén alapul, és azon, hogy kivételes hatás előrejelzése nem szükségképpen feltételezi rendkívüli képességek észlelését. Abból indulunk ki, hogy megérthetjük, kvantitatív módon leírhatjuk, és akár előre is jelezhetjük a sikert csupán a sikert vezérlő kollektív társadalmi jelenségek feltárása alapján.

Annak ellenére, hogy a probléma régóta az érdeklődés középpontjában áll, még szakemberek is rendszeresen tévednek a siker, a népszerűség és a hatás hosszú távú előrejelzésében. A siker azonban egyszer csak előre jelezhetővé válik, ha nem egyéni, hanem kollektív tevékenységnek tekintjük: ahhoz, hogy valami sikeres legyen, nem elég újnak és vonzóknak lennie, hanem széles körű egyetértésre van szükség abban, hogy dicséretre érdemes. Ha elfogadjuk a siker kollektív természetét, a hálózat- és adattudomány segítségével fel lehet tárni kvantitatív jellemzőit a számos rendelkezésre álló adatból (Clauzet et al., 2017).

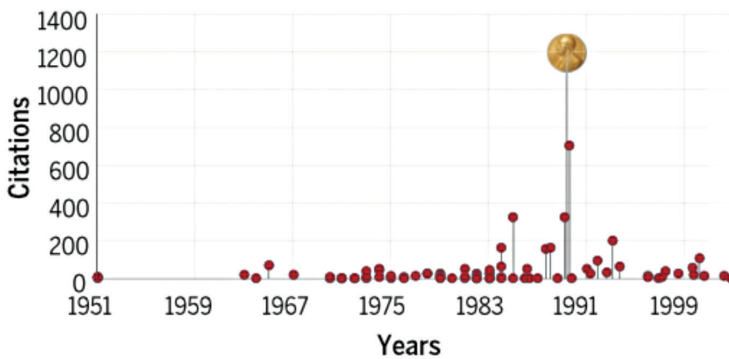
A teljesítmény és a siker dichotómiájának eredete az, hogy az emberi tevékenységek széles körében nehéz megkülönböztetni a teljesítményt és a sikert (Murray, 2003). Valóban, idézetséget, díjakat és más mértékeket használunk egy tudományos felfedezés értékének mérésére, ugyanakkor hiányoznak a teljesítmény objektív mérőszámai, amelyek egy tudományos dolgozat vagy tudós innovációs mértékét vagy tehetségét jellemeznék. Kutatásainkban elsősorban azon mechanizmusok kvantitatív leírására összpontosítottunk, amelyek a tudományos munka sikerét hajtják, mivel ezen a területen kitűnő példákat találunk azon kérdésekre, amelyeket a siker szociológiája meg tud válaszolni.

Egyik legutóbbi munkánkban az egyéni tudományos hatás fejlődését vizsgáltuk (Sinatra et al., 2016). Az emberi teljesítmények legtöbb területén, a sporttól a mérnökségig, a nagyobb eredményhez vezető út meredek tanulási szakaszból és hosszú gyakorlásból áll. A tudomány sem igazán különbözik: a kiemelkedő felfedezéseket gyakran kevésbé emlékezetes hatású dolgozatok előzik meg. Azonban annak ellenére, hogy egyre növekszik az igény az ígéretes fiatal tudósok korai azonosítására, a tudományos kiválóság kiemelkedésére jellemző pályafutás időbeli mintázatát nem sikerült felfedni. Hogyan változik a hatás és a produktivitás egy tudományos pályafutás során? Követ a hatás valamilyen előre jelezhető mintázatot? Előre tudjuk jelezni egy tudós kiemelkedő eredményének idejét? Képesek vagyunk tudományos pályafutásokat modellezni mérhető és előre jelezhető mennyiségek segítségével? Ezek a kérdések motiváltak, amikor elhatároztuk, hogy tudományos karrierük ezereinél számszerűsítjük a hatás és produktivitás fejlődését. Negyvenmillió tudományos közlemény és hivatkozásaik felhasználása tette lehetővé, hogy hét diszciplína tudósainak publikációs jegyzékét rekonstruáljuk, és azonosítsuk a dolgozatoknak a tudományos közösségre gyakorolt hosszú

távú hatását az idézettségi mutatókkal. Azt találtuk, hogy a legnagyobb hatású dolgozat egy tudós pályájának véletlenszerű időpontjában jelentkezik. Más szóval, a legnagyobb hatású munka egyforma valószínűséggel található a tudós dolgozatainak sorozatában – lehet az az első publikáció, felbukkanhat a pályafutás közepén, vagy lehet a tudós utolsó dolgozata (5. ábra).



Frank A. Wilczek  
Physics Nobel,  
2004



John B. Fenn  
Chemistry Nobel,  
2002

**5. ábra.** A véletlen hatás szabály. Két Nobel-díjas tudós, Frank A. Wilczek (fizikai Nobel-díj, 2004) és John B. Fenn (kémiai Nobel-díj, 2002) publikációinak története illusztrálja, hogy a legnagyobb hatású dolgozat azonos valószínűséggel lehet akárhol a tudós publikációinak sorozatában. A függőleges vonalak egy-egy tudományos cikknek felelnek meg. A vonalak hossza a cikk hatását mutatja, vagyis azt, hogy hány hivatkozást kapott a cikk tíz év után. Wilczek a Nobel-díjat a legelső publikált dolgozatáért kapta, míg Fenn a Nobel-díjjal jutalmazott cikkét karrierje kései szakaszában írta, azután, hogy nyugdíjba küldték a Yale-ről.

(Az ábra átvétel a Sinatra et al., 2016 cikkből)

Ez a véletlen hatás szabály érvényes különböző diszciplínák tudósaira, akik különböző hosszúságú életpályákat futottak be, különböző évtizedekben dolgoztak, és egyedül vagy teamben publikáltak. Ez a szabály lehetővé tette, hogy kifejlesszünk egy kvantitatív modellt, amely szisztematikusan szétválasztja a produkti-

vitás és a szerencse szerepét a tudományos karrierben. A modell minden tudóshoz egyértelműen hozzárendel egy paramétert, ami leírja a tudós képességét arra, hogy nagy hatású dolgot hozzon létre, és független a tudós életpályájának szakaszaitól. A modell analitikus kifejezéseket szolgáltat a hatás szokásos mértékeire, így a *h*-indexre, vagy a hivatkozások teljes számára, és lehetővé teszi számunkra, hogy előre jelezzük az egyes tudósok időbeli fejlődését, beleértve a független elismeréseket, mint a Nobel-díj előrejelzését is (Sinatra et al., 2016). Mindez együtt arra vezet, hogy nem tudjuk megjósolni, hogy egy karrieren belül *mikor* születik meg a legnagyobb hatású cikk, de előre tudjuk jelezni, hogy *kinek* lesz nagy hatása, és mi lesz az egyes pályafutások hatásának a fejlődése. A „szerencse” és az egyéni képességek szétválasztása, ami az egyes pályafutásokon mérhető, lehetővé teszi, hogy minden tudósra számszerűen meg tudjuk becsülni a legnagyobb hatású munkájának az *impaktját* és az *impakt* indikátorok dinamikáját akár évtizedekre előre.

Néhány további példa a CEU-s csoport további eredményeiből a siker szociológiájának területén: dolgozatok sikerességének hosszú távú, akár harminc évre vonatkozó előrejelzése (Wang et al., 2013), az elismerés allokálása egyes tudóshoz többszerzős cikkeknel (Shen–Barabási, 2014) az interdiszciplinaritás számszerűsítése a hatás alapján és kapcsolata tekintélyes díjakkal, mint a Nobel-díjjal (Szell–Sinatra, 2017), valamint a teljesítmény és a népszerűség kapcsolatának megértése egyes esetekben, amikor objektív mérték áll rendelkezésre mindkettőre, mint a sportban (Yucesoy–Barabási, 2016).

Mindezen megközelítések szigorú, kvantitatív keretet adnak a siker számszerűsítéséhez és előrejelzéséhez. Az így elért megértés segíthet a teljesítmény jobb megítélésében, és hozzájárulhat, hogy felismerjük a korlátait annak, ha az emberi teljesítményt a siker alapján akarjuk számszerűen megítélni. Hozzájárulhat ahhoz is, hogy korai életszakaszukban azonosítani lehessen személyeket, akik nagy teljesítmények elérésére képesek, és tanácsokkal szolgáljon a döntéshozóknak a támogatásukhoz.

## ÖSSZEFOGLALÁS

A fenti példákkal betekintést kívántunk nyújtani azon kutatások egy részébe, amelyek a számítógépes társadalomtudomány keretében a CEU-n folynak. Az ötletek sikerre vitelének különböző szempontjait próbáltuk megvilágítani, a kreatív csapatok kialakítási feltételeitől az innováció terjedési mechanizmusának leírásán keresztül a siker elméletének a tudományos eredményekre történő alkalmazásáig. Nem állítjuk, hogy máris sikerült egységes képbe foglalni ezt a szerteágazó és fontos témakört, de az adatok feldolgozása egészen újszerű megközelítéseket tett lehetővé, illetve közvetlenül tesztelhetővé tette az elméleti modellezést. Ezeket az előnyöket kihasználva, a jövőben további adatokat bevonva tervezzük folytatni a

kutatásokat. A cél, hogy kvantitatív elméleti leírást tudjunk adni a folyamatokra, így lehetővé téve az előrejelzést, aminek óriási gazdasági-társadalmi haszna lehet.

### KÖSZÖNETNYILVÁNÍTÁS

A kutatást az OTKA (K112713), továbbá a H2020, az AFOSR és az NSF támogatta.

### IRODALOM

- Aral, S. – Van Alstyne, M. (2011): The Diversity-Bandwidth Trade-Off. *American Journal of Sociology*, 117, 90–171. DOI: 10.1086/661238, [http://www.jstor.org/stable/10.1086/661238?seq=1#page\\_scan\\_tab\\_contents](http://www.jstor.org/stable/10.1086/661238?seq=1#page_scan_tab_contents)
- Barabási A-L. – Albert R. (1999): Emergence of Scaling in Random Networks. *Science*, 286, 509–512. DOI: 10.1126/science.286.5439.509, <http://barabasi.com/f/67.pdf>
- Chovell, G. et al. (2017): Perspectives on Model Forecasts of the 2014–2015 Ebola Epidemic in West Africa: Lessons and the Way Forward. *BMC Medicine*, 15, 42, DOI: 10.1186/s12916-017-0811-y, <https://bmcmmedicine.biomedcentral.com/track/pdf/10.1186/s12916-017-0811-y>
- Clauset, A. – Larremore, D. B. – Sinatra, R. (2017): Data-driven Predictions in the Science of Science. *Science*, 355, 477–480. DOI: 10.1126/science.aal4217, <http://science.sciencemag.org/content/355/6324/477.full>
- De Vaan, M. – Stark, D. – Vedres B. (2015): Game Changer: The Topology of Creativity. *American Journal of Sociology*, 120, 4, 1144–1194. [http://www.personal.ceu.hu/staff/Balazs\\_Vedres/papers/devaan.vedres.stark.games.pdf](http://www.personal.ceu.hu/staff/Balazs_Vedres/papers/devaan.vedres.stark.games.pdf)
- Granovetter, M. (1973): The Strength of Weak Ties. *American Journal of Sociology*, 78, 6, 1360–1380. DOI: 10.1086/225469, [https://www.jstor.org/stable/2776392?seq=1#page\\_scan\\_tab\\_contents](https://www.jstor.org/stable/2776392?seq=1#page_scan_tab_contents)
- Karsai M. – Iniguez, G. – Kikas, R. et al. (2016): Local Cascades Induced Global Contagion: How Heterogeneous Thresholds, Exogenous Effects and Unconcerned Behavior Govern Online Adoption Spreading. *Science Reports*, 6, 27178. DOI: 10.1038/srep27178, <https://www.nature.com/articles/srep27178>
- Ke, Q. – Ferrara, E. – Radicchi, F. – Flammini, A. (2015): Defining and Identifying Sleeping Beauties in Science. *Proceedings of the National Academy of Sciences of the USA*, 112, 7426–7431. DOI: 10.1073/pnas.1424329112, <http://www.pnas.org/content/112/24/7426>
- Lord, T. (2010): *The Jazz Discography*. <https://www.lordisco.com/>
- Murray, C. (2003): *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 BC to 1950*. New York: Harper Collins
- Obstfeld, D. (2005): Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation. *Administrative Science Quarterly*, 50, 100–130. DOI: 10.2189/asqu.2005.50.1.100, [https://www.jstor.org/stable/30037177?seq=1#page\\_scan\\_tab\\_contents](https://www.jstor.org/stable/30037177?seq=1#page_scan_tab_contents)
- Ruan, Z. – Iniguez, G. – Karsai M. – Kertész J. (2015): Kinetics of Social Contagion. *Physical Review Letters*, 115, 218702 DOI: 10.1103/PhysRevLett.115.218702, <https://arxiv.org/pdf/1506.00251.pdf>
- Sinatra, R. – Wang, D. – Deville, P. et al. (2016): Quantifying the Evolution of Individual Scientific Impact. *Science*, 354, aaf5239, DOI: 10.1126/science.aaf5239, [https://www.researchgate.net/publication/309692079\\_Quantifying\\_the\\_evolution\\_of\\_individual\\_scientific\\_impact](https://www.researchgate.net/publication/309692079_Quantifying_the_evolution_of_individual_scientific_impact)

- Shen, H. W. – Barabási A-L. (2014): Collective Credit Allocation in Science. *Proceedings of the National Academy of Sciences of the USA*, 111, 34, 12325–12330. DOI: 10.1073/pnas.1401992111, <http://www.pnas.org/content/111/34/12325>
- Szell, M. – Ma, Y. – Sinatra, R. (2018): Interdisciplinarity: A Nobel Opportunity. (Preprint)
- Török J. – Kertész J. (2017): Cascading Collapse of Online Social Networks. *Science Reports*. 7, 16743. DOI:10.1038/s41598-017-17135-1, <https://www.nature.com/articles/s41598-017-17135-1.pdf>
- Uzzi, B. (2008): A Social Network's Changing Statistical Properties and the Quality of Human Innovation. *Journal of Physics A: Mathematical and Theoretical*, 41, 224023. DOI: 10.1088/1751-8113/41/22/224023, <https://bit.ly/2KzwPj9>
- Vedres B. – Stark D. (2010): Structural Folds: Generative Disruption in Overlapping Groups. *American Journal of Sociology*, 115, 4, DOI: 10.1086/649497, <http://www.jstor.org/stable/pdf/10.1086/649497.pdf?refreqid=excelsior%3A99619439d35b35221326a91c4a44f95f>
- Vedres B. (2017): Forbidden Triads and Creative Success in Jazz: The Miles Davis Factor. *Applied Network Science*, 2, 31, <https://arxiv.org/ftp/arxiv/papers/1709/1709.03787.pdf>
- Wang, D. – Song, C. – Barabási A-L. (2013): Quantifying Long-term Scientific Impact. *Science*, 342, 127–132. DOI: 10.1126/science.1237825. <http://science.sciencemag.org/content/342/6154/127.full>
- Watts, D. (2002): A Simple Model of Global Cascades on Random Networks. *Proceedings of the National Academy of Sciences of the USA*, 99, 5766–5771. DOI: 10.1073/pnas.082090499, <http://www.pnas.org/content/99/9/5766>
- Watts, D. – Strogatz, S. H. (1998): Collective Dynamics of 'Small-world' Networks. *Nature*, 393, 440–442. DOI:10.1038/30918, [https://www.researchgate.net/publication/13660517\\_Collective\\_Dynamics\\_of\\_Small\\_World\\_Networks](https://www.researchgate.net/publication/13660517_Collective_Dynamics_of_Small_World_Networks)
- Yucesoy, B. – Barabási A-L. (2016): Untangling Performance from Success. *EPJ Data Science*, 5, 17, <https://link.springer.com/article/10.1140/epjds/s13688-016-0079-z>