

Digitális Pareto – az online közösségeken belüli megoszlás mintázatairól

A digitális információk mennyisége robbanásszerű növekszik. A digitális lábnyomok eddig megfigyelhetetlen viszonyokat, hálózatokat tesznek láthatóvá és kutathatóvá. Ennek nyomán az online közösségek szerkezetére vonatkozóan két elméleti pozíció kristályosodik ki skála független megoszlások és a long tail elmélet. Cikkünkben amellet érvelünk, hogy ez egy empirikusan eldöntendő kérdés, és a következő időszak egy fontos kutatási iránya, kezdő lépése lesz az egyes közösségek estében.

Kulcsszavak: *digitális lábnyomok, online közösségek, Pareto, power law, struktúra, megoszlás, Long Tail, skálafüggetlen megoszlások*

Szerzői információ:

Dessewffy Tibor szociológus, (ELTE TÁTK) a Digitális Szociológia Központ igazgatója, a Szociálpszichológia tanszék vezetője. Kutatási területei a Big Data, digitális kultúra és az információs társadalom.

Láng László a közgazdaság tudományok kandidátusa, az International Business School rektora. Érdeklődési területei a Big Data, a fejlődésemeltek és a globális egyenlőtlenségek. Főbb publikációi Dessewffy Tiborral: Big Data és a társadalomtudományok véletlen találkozása a mítóasztalon (Replika 92-93.), Kód(a) (Replika 92-93.), Egyetem 2.0? Felsőoktatás az információs korban (Magyar Tudomány 2015/5.).

Így hivatkozzon erre a cikkre:

Dessewffy Tibor, „Digitális Pareto – az online közösségeken belüli megoszlás mintázatairól”.

Információs Társadalom XV, 3. szám (2015): 18–40.

<https://dx.doi.org/10.22503/inftars.XV.2015.3.1>

A folyóiratban közölt művek

a Creative Commons Nevezd meg! – Ne add el! – Így add tovább! 4.0

Nemzetközi Licenc feltételeinek megfelelően használhatók.

Dessewffy Tibor – Láng László

Digitális Pareto – az online közösségeken belüli megoszlás mintázatairól

Bevezetés: Big Data, mint látásmód

Napjaink egyik legdrámaibb folyamata az információs robbanás és az ennek nyomán végbemenő drámai társadalmi átalakulás. Cukier és Mayer-Schönberger könyvének címevel: *Big Data, a forradalom, amely megváltoztatja az életünket, a munkánkat és a gondolkodásunkat* (Cukier és Mayer-Schönberger, 2013). De nem csak arról van szó, hogy a Big Data nyomán az egyes életszférák, foglalkozások, tevékenységformák változnak a kereskedelemtől a gyógyászatig, a repülőgép szereléstől a párválasztásig. Hanem ezzel egyidejűleg módosul a megismerés eszköztára, módszertana és eredményei is (Dessewffy-Láng, 2015; Csepeli-Dessewffy, 2015). Ahogy Latour is hangsúlyozza: „*Ha megváltoztatjuk a megfigyelés eszközeit, megváltozik az egész társadalomelmélet*” (Latour, 2010, 155. o.). Ennek az episztemológiai jellegű fordulatnak az egyik legfontosabb eleme, hogy a korábban egységesnek, homogénnek tételezett holisztikus fogalmak felbomlanak. A digitális lábnyomokból rekonstruálhatóak eddig rejtett szerkezetek és összefüggések. Ahogy a mikroszkóp feltalálása lehetővé tette sejtek és mikroorganizmusok felismerését és ezáltal az emberi test működésének pontosabb megértését, a Big Data alapú megközelítések is segítenek abban, hogy a korábban láthatatlan szereplőket és köztük levő viszonyrendszereket feltárjuk.

Társadalomtudományi példával élve: a Big Data megadja a lehetőséget Moreno életművének kiteljesítéséhez. Morenót az a probléma foglalkoztatta, hogy az egyéni viszonyokból felépülő csoportok, szervezetek és társadalmak működését alaposabban megértse. Ráébredt arra, hogy az egyes csoportokat jelölő címkék mögé kell néznie, azokat a szerkezeteket kell megismerni, amelyek meghatározzák a közösségek sikeres működést. Itt nem pusztán egyéneket látunk, hanem egyének közötti dinamikus viszonyokat. Morenónál a „társadalmi atom” nem az izolált egyén, hanem a kapcsolatok és remélt kapcsolatok „elektronjai”-nak mozgáspályái által megragadható egyén (Moreno, 1947). E viszonyok megragadásra született meg a szociometria, amely a közösségek rejtett hálózatát kívánta feltárni. A fiúkat és lányokat jelképező háromszögeket és négyszögeket összekötő szimpátiát és ellenszenvet kifejező fekete és piros vonalakkal előállított szociogramok ma talán már nem tűnnek túlságosan izgalmasnak. De ha, mint a térkép metaforájaként kezeljük őket, mely térképek eligazítanak az addig ismeretlen tájakon, segítenek célba érni, ezáltal új utak és tevékenységek számára nyitnak teret, nos, akkor változatlanul nagyjelentőségű eszkről és megközelítésről van szó. Moreno ambíciói messze túlmutattak a kicsocportok viszonyainak feltérképezésen, egész városokban és nemzetekben gondolkodott, vagy ahogy fogalmazott, az emberiség egésze érdekelte. De érezte azt is, hogy a korabeli eszközök, statisztikai apparátusok kemény korlátot szabtak ambíciójának. „*Abszolút fontos lenne, hogy ismerjük az emberi társadalmak szerkezetét egy adott pillanatban. Ám hatalmas és riasztó nehézségekbe ütközünk, midőn erre a törekszünk. E nehézségek lényegében három kategóriába sorolhatóak: az emberek nagy száma, a tényleges részvétel mérésének szükségessége és a folyamatos és megismételt kutatások szükségessége.*” (Moreno, 1937, 206. o.) Moreno megközelítésének a



lényege, hogy a személyközi viszonyokat nem szimbolikus fogalmak használatával, hanem a maguk konkrétságában kell vizsgálni, lehetőleg a szereplők spontán viselkedést és nem torzító önreprezentációit alapul véve. *„Minden személyt a maga konkrétságában és nem, mint valami szimbólumot kell vizsgálnunk, és a személyek közötti viszonyokat is a maguk konkrétságában kell vizsgálnunk. Tudásunk mindaddig nem lehet teljes, amíg mindenki spontánul és legjobb képessége szerint részt nem vesz e viszonyok feltárásában, bemutatásában.”* (Moreno, 1937, 207. o.)

A Big Data új korszaka megteremti a lehetőséget ezeknek az elvárásoknak a beteljesítésre, hiszen akár nagyszámú csoport spontán viselkedéséről szerezhethünk korábban elképzelhetetlen mélységben konkrét információkat.¹ Christian Rudder, a Harvard Egyetem volt matematikus hallgatója három társával megalapította az OKCupid ismerkedő oldalt. Az oldalnak évente 11 millió aktív használója van, akiknek preferenciáit, választásait és online viselkedését, önprezentációit és választásait az OKCupid ismeri és elemzi. Az adatbányászat révén lehet a cég sikeres: olyan tagokat tud egymásnak ajánlani, akik folyamatosan tökéletesített algoritmusai szerint jól passzolnak egymáshoz. Ezt az adatvagyon nem csak az ajánlórendszer hatékonyságának növelésére lehet használni, hanem ahogy Rudder Dataclysm című könyvéből kiderül, számos korábban nem bizonyított összefüggés bemutatására (Rudder, 2014). (Hogy csak egyet említsünk ilyen a férfiak nő ideális korára vonatkozó kijelentéseik, valós keresési magatartásuk és konkrét választásaik közötti diszcrepancia.) Rudder olyan adatbázist – és részben elemzést – alkotott, amely a maga konkrétságával, nagyságával, többretegű kapcsolat elemzésével és időbeli dinamizmusával kiteljesítette és megvalósította Moreno eredeti programját.

Ebben a tanulmányban a mi ambícióink szerényebb. A fentiek alapján abból indulunk ki, hogy a digitális adatrobbanás nyomán előálló Big Data korszakban eddig nem ismert komplexumát ismerhetjük meg a szereplőknek és az őket összekapcsoló folyamatoknak. Ez esetben a hálózati struktúrák feltárása előtt is adódik egy kérdés, nevezetesen hogy a vizsgált közösségek milyen megoszlásokat mutatnak egy adott viszonyban. Mint látni fogjuk, e kérdés jelenleg komoly viták középpontjában áll – és a megoldás nem triviális. Az alábbiakban előbb áttekintjük a megoszlásokkal kapcsolatos konkrét dilemmát: a Long Tail–Long Neck vitát. Majd különböző példákon keresztül azt próbáljuk bizonyítani, hogy az internetes világban e megoszlások a kontextustól függően különbözőek lehetnek. (Az a szabály hogy nincs szabály). Végül saját empirikus adatainkon keresztül érvelünk, e kontextusok egy viszonylag elhanyagolt tényezője, a zsánerek figyelembe vétele mellett.

Long Tail vagy Long Neck dilemma

A paradigmaváltásra vonatkozó bármely igény csak akkor tekinthető indokoltnak, ha a valóság olyan szeleteit is képes megmagyarázni, amely a korábbi korszakban az eldönthetetlen kérdések körébe tartozott. Annak érdekében, hogy szemléltessük – ehelyütt inkább csak felvillantsuk, mintsem átfogóan elemezzük – a Big Data alapú megközelítésben rejlő lehetőségeket, vegyünk egy empirikus kérdést, a „Long Tail” vs. „Long Neck” dilem-

¹ Jelen gondolatmenetünknek nem része, hogy részletesen tárgyaljuk a Big Data „sötét oldalát”. Itt csak megemlítjük, hogy ez a folyamat rendkívül komoly személyiségszociológiai, biztonsági, társadalmi kérdéseket is felvet.

máját. Korántsem marginális kérdésről van szó: a dilemma kulturális fogyasztástól a társadalmi struktúra elméletekig napjainkban mindenütt megjelenik.

Chris Anderson eredeti, Long Tail című írása főként az online kereskedelem sajátosságaival, az ebben rejlő lehetőségekkel foglalkozott (Anderson, 2004). A cikk és a később ebből születő könyv elsöprő sikerének alighanem az az oka, hogy Anderson téziseinek *messzire vezető társadalmi hatásai, illetve tágabb értelmezési lehetőségei is vannak* (Anderson, 2006). A „Long Tail” abból indul ki, hogy az internet világában a Pareto-elv 80/20 megoszlása nem működik.² A Pareto-elv szerint a hagyományos kisszámú termék generálja a bevételek túlnyomó többségét. Ezzel szemben a digitális világban a kasszasikerek, az ikonok nem tűnnek el, de felértékelődik a hosszú, vízszintesen elnyúló alagútban a kis számosságú választások szerepe. Az eredeti 80/20 arány értelmezhetetlenné válik, mert a korábban zéróösszegű felosztás a szinte végtelen választási lehetőségek esetén alapvetően átalakul. Egy hagyományos könyvesbolt szükségszerűen korlátos polcain érvényesülhetett a Pareto-elv, de ez, érvel Anderson a digitális térben, például az Amazon esetében már nem működik. Az Amazon ugyanis elvileg végtelen számú könyvet kínál olvasóinak – ezáltal megfordítva a hagyományos képletet, ahol a könyvesboltban végül is csak a tulajdonos, a kereskedők által szelektált készletből vásárolhatunk. Mivel alighanem mindenki ismeri azt az élményt, hogy mennyire kicsinek is érezheti magát az ember egy hatalmas könyvesboltban, azt is gondolhatnánk, hogy ez a kínálat éppen elég, hiszen már így is sokkal több könyv van a polcokon, mint amennyit képesek lennénk elolvasni. A Long Tail-elmélet másik fontos tétele épp erre figyelmeztet: az online térben nincs olyan nagy mennyiségű kínálat, amire ne lenne valamekkora igény. És ez több mint az egzotikus kuriózumok iránti esetleges érdeklődés. Nem pusztán azt állítjuk tehát, hogy a Long Tailben található termékek iránt is van minimális érdeklődés, hanem azt, hogy az itt található termékek hatalmas mennyisége miatt ez üzletileg, az összkép szempontjából is jelentős. Az Amazon esetében például, Anderson számításai szerint a könyveladási bevételek 57%-a olyan könyvekből származik, amelyek nem elérhetőek a hagyományos könyvesboltokban.

A Long Tailen belüli információáramlás és információcsere lehetőségét az ajánlórendszerek, peer-to-peer, vagyis felhasználtól felhasználóig vezető algoritmusai teremtik meg. Ha valaki norvég trash metált keres az interneten, szinte nem is tudja elkerülni, hogy felajánlják neki, hogy az e szubkultúra iránt rajongók által meghallgatott további számokat megismerje. Ez könyvekre, filmekre, videókra, blogokra, újságcikkekre egyaránt igaz.

Szociológiai értelemben ennek az új üzleti modellnek messze ható következményei vannak. Nevezetesen, a digitális korban a Long Tailben a kulturális fogyasztás sajátos mintázatai mentén szerveződő identitások nem csak ellensúlyozzák a nagy erőforrásokkal felújított kasszasikerek mentén szerveződő fogyasztási, életstílus és identitás csoportokat, hanem hosszabb távon elkerülhetetlenül háttérbe is szorítják azokat.

Anderson elmélettel szemben is jól definiálható egy teoretikus pozíció: Barabási Albert László az internetstruktúráját tekinti skálafüggetlennek, ahol a csomópontok szerepe a sokszorosára növekszik (Barabási, 2002). Anita Elberse (2013) a kulturális fogyasztás terü-

² Vilfredo Pareto tétele szerint a következmények 80% az okok 20%-ra vezethető vissza. Érdemes megjegyezni, hogy bár Pareto a kertjében növő zöldborsó termésmegoszlásának megfigyeléséből indult ki – vagyis a termés 80% a borsóhévelyek 20%-ból származik –, de általános szabályként tekintett tételére, amelyet a földbirtok megoszlástól a vagyommegoszlásig alkalmazott a legkülönbözőbb területeken.

letén érvel az elsőrő kasszasikerek változatlan dominanciája mellett. A közgazdász Tyler Cowen az átlag eltűnését, a digitalizáció és a mesterséges intelligencia elterjedése miatt a középosztály elolvadását, a top 15-20% további gazdasági és politikai erősödését diagnosztizálja (Cowen, 2013). Azonos ezekben a megközelítésekben, hogy ezek a kutatók is kiemelkedő jelentőséget tulajdonítanak a digitális változásoknak. Ám míg Andersen a Long Tailben zajló folyamatokat és ezek komplexitásának növekedését látja legfontosabbnak, addig Barabási, Elberse, Cowen és mások a skálafüggetlen csomópontok, a kasszasikerek, a domináns elit fogalmait hangsúlyozzák. Ezek a megközelítések „Long Neck” elméletekként összegezhetők. A Long Neckben nem a vízszintes tengely mentén borul a Pareto-elv, nem a potenciális résztvevők száma növekszik, hanem a függőleges tengely mentén a mindenkori kevesek gyarapodnak tovább.

Long Tail vs. Long Neck – empirikus illusztrációk

Ahhoz, hogy ezt a kérdést eldönthessük először is redukálnunk kell az itt vázolt komplexitást – bár a Long Tail elmélet számos területen inspiráló, választanunk kell a lehetséges kontextusok közül. Merész feltételezés lenne azt állítani – még ha egyes szerzők bevállalják is ezt –, hogy valamennyi területen azonos dinamika működik.

Vegyük először is az online felsőoktatás napjainkban kialakuló piacát. Ez első látásra a Long Tail logikát követi: ma már bárhol bárki feltöltheti kurzusait az Internetre. Ám lehet érvelni amellest is, hogy valójában „Long Neck” logika működik. Például az „Apple iTunes U platformján több mint 2500 kurzus fut, melyeknek több mint negyed millió hivatalos diákja van, a letöltések száma 2013-ban meghaladta az 1 milliárdot. Az olyan nagynevű egyetemek, mint a Yale, a Cambridge, az MIT vagy az Oxford tovább erősítik vezető pozíciójukat azért, hogy egyetlen iTunes U kurzusukra több mint százezren jelentkeznek. A Stanford University és a The Open University tananyagai egyaránt túllépték a 60 milliós letöltést. Az Ohio State Universityn Matthew „Dr. Fus” Stoltzfus Általános Kémia kurzusára a meghirdetés évében több mint 100 000 jelentkeztek.³ Ebből arra is következtethetünk, hogy bár az elérhető kurzusok száma folyamatosan növekszik, a hallgatók (a „vásárlók”) a korábban nagy nevet szerzett intézmények es előadók anyagait követik. A jelen gondolatmenet számára két dolog fontos:

Először is, hogy a Long Tail–Long Neck dilemma eldöntése tipikusan empirikus kérdés – hiszen mindkét hipotézis mellett lehet konzisztensen érvelni – hogy melyik bizonyul igaznak, azt csak méréssel lehet eldönteni. Másodsor az, hogy az egymilliárd hallgató hogyan oszlik meg, milyen aktivitást mutat, milyen jellegzetességekkel – ez tipikusan Big Data probléma. Abban az értelemben, ahogy fentebb definiáltuk – vagyis survey módszerekkel nem megválaszolható kérdés.

Vegyük példának a web 2.0-ás alkalmazások egyik legizgalmasabbját a Kickstartert. Mint ez közismert, a Kickstarteren a megjelenített ötletekhez keresnek anyagi támogatást, finanszírozást. Ehelyett nem szükséges a Kickstarter egyébként rendkívül izgalmas működését részletesen elemezni. A jelen érvelés szempontjából fontosabb az, hogy előzetesen, pusztán elméleti alapon eldönthetetlen, hogy a Kickstarteren megjelenő közel 1 milliárd dollár, hogyan oszlik meg. Konzisztensen érvelhetnénk amellest, hogy a Long Neck logikát követve, néhány kiemelkedő, világsikert hozó ötlet, mint a Peeble okos óra vagy az Ouya já-

³ <http://www.apple.com/pr/library/2013/02/28iTunes-U-Content-Tops-One-Billion-Downloads.html>

tékkonzol vonzza a legnagyobb tőkét, és ezek a meghatározóak a támogatások összességében. De a Long Tail elv érvényesülése mellett is hozhatóak fel ugyanilyen erős érvek, vagyis hogy nagy tömegű, de kis összegű támogatások a meghatározóak a finanszírozásban. Egy nemrégiben publikált empirikus kutatás segít eldönti ezt az elméleti kérdést.⁴ Amint az 1. ábrán is láthatjuk, a Kickstarteren 56 projekt volt képes 1 milliónál több támogatást szerezni. 1031 projektszervezet 100 000 és 1 millió dollár közötti támogatást. Ugyanakkor meghatározóak az alacsonyabb összegű támogatások.



1. ábra A Kickstarter projektek finanszírozási szintjeinek megoszlása

Összességében az 1000 legjobban támogatott projekt a Kickstarter 352 519 352 dollár értékben kapott támogatást, amely a bevételek 43 százalékát teszi ki. Ezen felül azonban további 56 035 támogatott projektet találunk, amely az ötletek 98%-t teszik ki. Ezek az ötletek százezer dollárnál kisebb támogatást kaptak, átlagosan 6743 dollárt. Összességében az ötletek gazdasági súlya, az általuk akkumulált tőke meghaladja a sikersztorik által vonzott befektetések nagyságát. Így ez a kutatás azt a korántsem triviális eredményt hozta, hogy a crowdsourcing alapú pénzügyítés logikája a Long Tail elvet követi.

A Long Tail-Long Neck megoszlások kérdése kapcsán valójában a görbe ívéről beszélünk. A jelen gondolatmenet kapcsán azt kívántuk bizonyítani, hogy e görbével pontos vonalának meghatározása empirikus kérdés, ahol már csak az adatok jellege miatt is, a Big Data módszertan megkerülhetetlen.

Fejtegetéseinket lezárva egy pillanatra visszatérnénk a társadalomelmélet szintjére. Ezt a témakört azzal indítottuk, hogy a hatványfüggvényekkel a társadalmi lét zavarba ejtően különböző területén találkozhatunk. Valószínűsíthető, hogy nincs egyetlen, mindenre illeszkedő görbe, amely ezt a sokszínűséget leírja. Ez nem különösebben megrázó állítás – ám nem

⁴ <http://qz.com/184019/how-kickstarter-users-raised-nearly-1-billion-the-really-long-tail-of-crowdfunding/>

nélkülözi az elméleti relevanciát. Castells trilógiájának is a Daniel Belltől kölcsönzött alap-gondolata, a „separate realm” tézise, vagyis, hogy a hatalom, a gazdaság és a kultúra területét más-más logikák és szerkezetek uralják (Castells, 2000; Bell, 1973). Azt a sejtést szeretnénk megfogalmazni, hogy az információs korban a gazdaság és a hatalom területén a Long Neck modell a domináns, vagyis egyre kevesebben rendelkeznek egyre nagyobb részesedéssel az e szférákra jellemző javakból. Ezzel szemben a digitális kultúra területén a szegmentáció, az egyénre szabottság és széttöredezettség a meghatározó-ezért itt inkább a Long Tail modell terjedését valószínűsítjük.

Long Tail vs. Long Neck – Online közösségek, prosumerek és zsánerek

Próbáljuk most meg az eddigi konfúziót némiképp redukálni. Brynjolfsson, Hu, és Simester arra hívják fel a figyelmet, hogy a Long Tail jelenség kapcsán érdemes megkülönböztetnünk a keresleti és a kínálati oldal folyamatait. A kínálat esetében az interneten eltűnik a polchelyek és bolti alapterületek szűk keresztmetszete, elvileg korlátlan mennyiségű áru jelenhet meg és ez már önmagában is produkálhatja a Long Tailt. Azáltal hogy nem 20 ezer, hanem kétmillió könyvből választhatunk, mint láttuk, megnő a vásárlásokban az eddigi elhanyagolt, boltokban nem kapható könyvek aránya (és csökken a szupersztárok súlya). A Long Tail *keresleti* oldalon ezzel szemben elsősorban az *információ áramlás* lehetősége teremti meg a niche termékek iránti igény megfogalmazására (Brynjolfsson, Hu, és Simester, 2011). Bár a Long Tail kínálati és keresleti oldalának megkülönböztetésnek érdekesnek tartjuk, ez a dichotómia figyelmen kívül hagyja az internet Web 2.0 korszakában végbemenő hibridizációt, kereslet és kínálat összekavaródását, a prosumerek (termelő/fogyasztók) megjelenését (Toeffler, 1980; Ritzer, Jurgenson, 2010). Az alábbiakban egy ilyen a prosumerek által meghatározott, Web 2.0 területet szeretnénk vizsgálni.

Konkréten az első látásra menthetetlenül Guttenberg Galaxishoz kötődő könnyolvasáshoz kapcsolódó két online közösséget. Habár az olvasásra fordított idő a különböző kutatások szerint csökken, az olvasás végéről szóló próféciák tévesnek bizonyulnak (Trend, 2010). Sőt, azt látjuk, hogy a Gigabyte Gestalt kitermel olyan kommunikációs mintázatokat, új közösségi formákat, amelyek tovább erősítik a könyvolvasás guttenbergi gyakorlatát. Ezek közül az egyik legjelentősebb a 2007-ben, Otis és Elisabeth Chandler által alapított Goodreads. A Goodreads, amelynek 2014-ben az USA-ban 21.5, globálisan 47.6 regisztrált tagja van a „könyvolvasás Facebookja”.⁵ A tagok nyilvánossá tehetik saját kedvenc könyveiket, az éppen olvasott műveket, láthatják barátaik, illetve kijelölt, követetni kívánt személyek kedvenceit, eddig és jelenleg olvasott könyveit. A social média prosumer logikáját követve, maguk is írhatnak kritikákat az egyes művekről, illetve pontozhatják azokat.⁶ 2014 júliusában több mint 34 millió a felhasználók által írt ilyen szöveges értékelést találhatunk a felületen.⁷ Itt a számok néhol valóban mellbevágóak: a „Harry Potter és a bölcsek köve” például a mai napon 2 821 817 pontozott értékelést, „rating”-et kapott az öt csillagból álló skálán, (melyek átlaga 4,38) az olvasók 43 080 szöveges kritikát írtak a könyvről.

A legnagyobb magyar közösségi könyves site-ot Nagy Bence újságíró alapította 2008-

⁵ (Could the Internet Save Book Reviews? Sarah Fay May 7 2012 The Atlantic <http://www.theatlantic.com/entertainment/archive/2012/05/could-the-internet-save-book-reviews/256802/>)

⁶ <http://www.theatlantic.com/entertainment/archive/2014/02/millions-of-people-reading-alone-together-the-rise-of-goodreads/283662/>

⁷ <http://www.statista.com/statistics/252998/number-of-user-generated-reviews-on-goodreadscom/>

ban. 2015-februárjában a moly.hu-nak közel 140 ezer tagja van, akik több mint 227 ezer könyvről 1 és egynegyed millió értékelést adtak.⁸ Az olvasói közösségben erős, moly.hu-s identitás alakult ki az oldalhoz köthető számos lehetséges tevékenység forma mentén. Ezek a könnykritikák írásától a kedvenc könyvek listázásán át a valódi könnycseréig, illetve az offline világban alakított moly csoportokig terjednek. A közösség a globális Goodreads-hez képest kisebb mérete ebben az esetben intenzívebb kapcsolatokat és erősebb identitást eredményezett (Dessewffy-Takáts, 2015). A Harry Potter könyvek itt is népszerűek, de az eltérő méreteket jól érzékelteti, hogy a „Bölcsek köve” 2915 csillagozott értékelést kapott, 92%-os tetszési mutatóval. A moly.hu-n is megjelennek más, leginkább angol nyelvű könyvek, így például a Harry Potter sorozat angol nyelvű nyitó darabja külön is feltűnik, (251 értékelést kap 95%-os tetszés mellett) a magyar nyelvű könyvek aránya elsöprő.

Visszatérve az előbbi Long Neck vs. Long Tail érveléshez: empirikus kérdés annak eldöntése, hogy ebben a térben milyen megoszlást követnek a könyvek értékelései. Empirikus kérdés, hiszen ahogy a Harry Potteres példa mutatja, érvelhetünk amellett, hogy a bestsellerek logikája érvényesül, vagyis néhány százaléknyi, nagy marketing befektetéssel hirdetett slágerkönyv dominálja a könyvfogyasztási piacot. Vagyis a paretoi 80/20 elv valamilyen arányban módosult verziójáról van szó, ahol a könyvek egy kis része kapja a beérkezett visszajelzések, kritikák nagy hányadát. De éppenséggel érvelhetünk ellenkezőleg is, hogy a nyilvánvalóan létező kasszasikerek ellenére jellemzőbb a könyvfogyasztásban az a sokszínűség, ahol a nagyszámú könyvre érkezett néhány értékelés a jellemzőbb, vagyis inkább a Long Tail modell érvényesül.

De lépünk tovább. Az eddigiekben világossá vált, hogy a Long Tail vagy a Long Neck különböző megoszlási szerkezeteket takarnak, amelyek eltérő terjedési mintázatok eredményei. A szokások, preferenciák és információk diffúziójával foglalkozó irodalom három szereplőt különböztet meg a terjedési folyamat sajátosságainak vizsgálatokor (Rogers, 2010). Először is a diffúziós aktorokat (véleményvezérek, összekötők, változás ágensek, korai adoptálók stb.) vagyis azok a szereplők, akik döntéseikkel és kommunikációjukkal alakítják a terjedési folyamatot (Watts, 2004; Kossinets és Watts, 2006). Másodszor a terjedési közeg sajátosságait vizsgálják: idetartozik a hálózatok szerkezete, a terjedést befolyásoló körülmények, a makro is mikro szintű gátló vagy serkentő tényezők. (Burt, 2000; Fowler – Christakis, 2010) Míg végül a harmadik lehetséges megközelítés a terjedés tárgyából indul ki, vagyis azt vizsgálja, hogy mi az, ami terjed, illetve magának a dolognak a sajátosságai mennyiben befolyásolják az elosztásokat. Általánosságban azt mondhatjuk, hogy míg az első két dimenzióban kimerítő vizsgálatokat folytatnak a társadalomtudományok különböző ágazatai, addig ez a harmadik megközelítés viszonylag alulreprezentált. Jóllehet Rogers klasszikus munkájában érinti az immanens tulajdonságok kérdését, de az általa feltett dimenziók a kipróbálhatóság, a korábbi rendszerekkel való kompatibilitás, az ár, a megfigyelhetőség és a kezelés könnyű megtanulása, elsősorban a technológiai eszközök terjedése esetében értelmezhető. (Rogers, 2010). Ez nem meglepő, hiszen az olyan „intellektuális technológiák”, mint a könyv vagy a zene ezek immanens vonzerejének vizsgálata – vagyis annak vizsgálata, hogy mennyire jó egy zeneszám vagy egy írásmű, rendszerint kívül esik a társadalomtudományok érdeklődésén (Carr, 2013). Esztétikai elemzésre mi sem vállalkozhatunk. De érdemes megvizsgálni, hogy az egyes zsánerek, illetve tematikák eltérő sajátosságokat mutatnak-e?

A zsáner repertoár fogalmát Orlikowski és Yates vezették be, a szervezeteken belüli kommunikáció különböző típusainak lehatárolásakor. Megítélésünk szerint a terminus azonban tágabb kontextusban is értelmezhető. Konkrétan a műfaji eltérések felől közelítve a fentebb

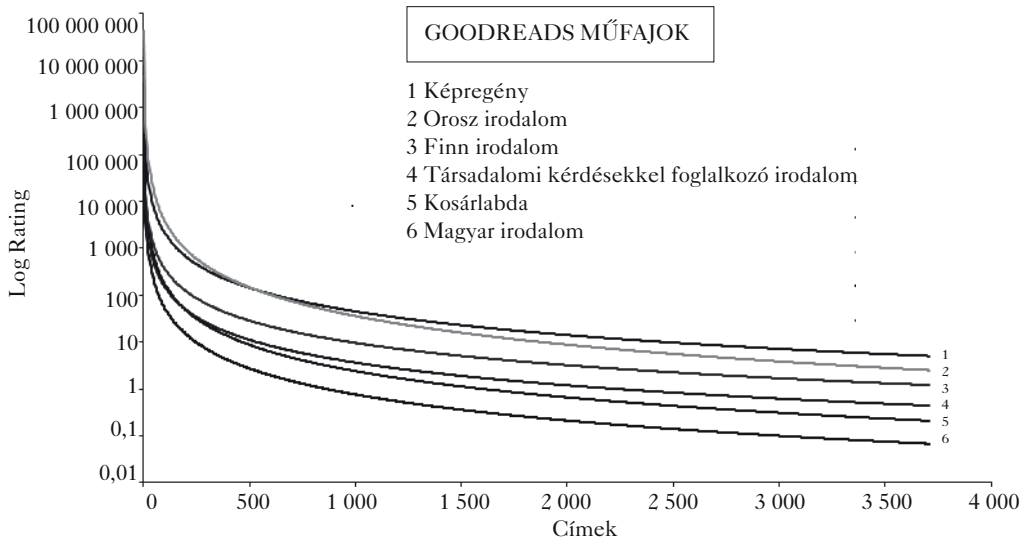
⁸ <http://moly.hu/rolunk> letöltve 2015. február 24.

tárgyalt kérdést, megvizsgálhatjuk, hogy az egyes zsánerek között vannak-e eltérések a megoszlások tekintetében.

A Goodreads-en és a moly.hu-n is tematikus bontásban is elérhetők a könyvek – így lehetőségünk van megnézni azt is, hogy az egyes zsánerekre milyen megoszlások jellemzőek. Itt a hipotézis megfogalmazás még nehezebb: ki tudná megmondani, hogy például a társadalmi mozgalmakról szóló könyvek értékelési görbéje, hogyan viszonyul a szakácskönyvek megoszlásához? Más szavakkal sztárok és bestsellerek a sikeres kevesek által uralt nyilvános tér jellemzőbb-e a szociológián belül, mint a szakácskönyvek világában?

Brynjolfsson, Hu és Simester cikkének egy másik fontos felismerése a koncentráció mérésének problematikája. Lényegében arról van szó, hogy az x tengelyen feltüntetett nagyon kis értékek figyelembe vételével előállíthatjuk a Long Tailt akkor is, ha ez valóságban nem létezik. Tipikusan ilyen a lurkerek problémája. Ha egy online fórum résztvevőinek aktivitást figyeljük, vagyis a rendszeresen (hetente) posztolókat, akkor azt találhatjuk, hogy a 100 résztvevő aktivitása jelentősen különbözik. Közelebről a posztolók feléhez kapcsolható a bejegyzések 75%. Amennyiben azonban figyelembe vesszük a site-ra regisztrált, de bejegyzéseket nem tevő *lurkerek, leskelődők* százas csoportjait is, akkor az a vizsgált csoport fele már az aktivitás 100% felel - miközben a valóságban nyilván nincs különbség (csak kérdésfeltevésben illetve a populáció megválasztásában). Ezt a problémát mi úgy kerüljük el, hogy nem a műfajokhoz tartozó könyvek olvasóit vizsgáljuk általában, mivel róluk nincsenek adataink, hanem csak azokat, akik a könyvekkel aktív viszonyba kerülve értékelték (Rating) vagy véleményezték (Review) azokat.

Anderson elmélete alapján okkal gondolhatjuk, hogy a könyvek értékelésében is a Long Tail megoszlással találkozunk: hiszen nincsenek fizikai, mennyiségi korlátok, a döntéseket meghatározó információ szabadon áramlik. Ám, mint ez az alábbi adatokból jól látszik, messze nem erről van szó, sokkal inkább a Pareto-szabály érvényesül, ahol a könyvek 20%-ához az aktivitás több mint 80% kapcsolódik. A Goodreads esetében a tages alapján hat olyan zsánert választottunk, ahol az elemszámok 4-5 ezer között mozogtak (2. ábra és 1. táblázat).

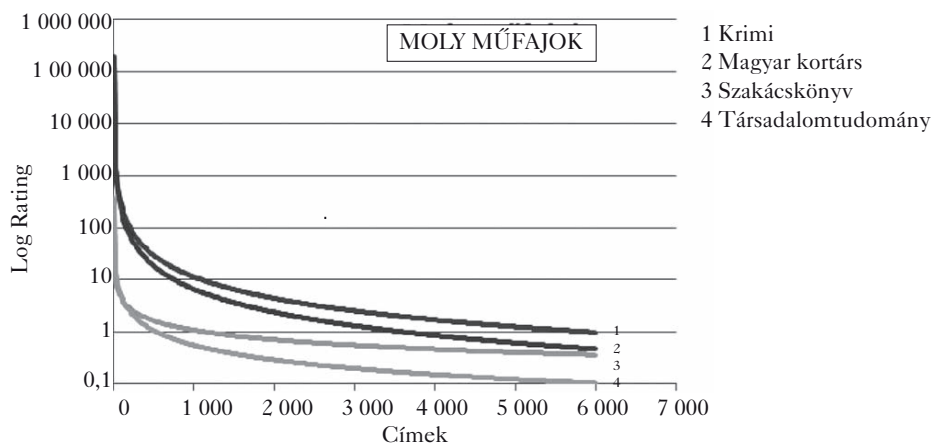


2. ábra Goodreads műfajok megoszlása

	Kosárlabda	Képregény	Finn irodalom	Orosz irodalom	Magyar irodalom	Társadalmi mozgalmak
Beszámolók						
10%	96,78%	86,49%	81,80%	93,30%	93,78%	87,26%
20%	99,07%	92,23%	88,78%	96,80%	96,37%	94,57%
Értékelés						
10%	94,81%	76,95%	78,56%	87,71%	90,55%	85,87%
20%	99,42%	87,60%	87,82%	94,24%	95,91%	94,71%

1. táblázat A beszámolók és az értékelések megoszlása a Goodreads rendszerében

Az individuális tagek a következők voltak: kosárlabda, képregény, finn irodalom, orosz irodalom, magyar irodalom, társadalmi mozgalmak. Mint láthatjuk, az egyes zsánerek között vannak különbségek, ám ezek elemzése kívül esik jelen tanulmány gondolatmenetén. Ami számunkra fontosabb, hogy a Goodreads értékelések estében egyértelműen a Long Neck modell érvényesül – hasonlóan a moly.hu-s értékelésekhez. A moly.hu estében az alábbi négy taggel jelölt, a hasonló elemszámú kategóriák megoszlását vizsgáltuk: társadalomtudomány, szakácskönyv, magyar kortárs irodalom, krimi. Bár a Goodreadsnél némiképp gyengébb mértékben, de itt is a Pareto-megoszlást találtuk, nem pedig a Long Tailt (3. ábra és 2. táblázat).



3. ábra A Moly.hu műfajok megoszlása

	Társadalom tudomány	Szakács könyv	Magyar kortárs irodalom	Krimi
Beszámoló				
10% (egész)	77,74	84,02	72,12	76,86
20% (egész)	91,42	100	86,53	87,22
Értékelés				
10% (egész)	78,73	73,68	72,2	76,64
20% (egész)	90,11	86,55	86,5	88,75

2. táblázat A beszámolók és az értékelések megoszlása a Moly.hu rendszerében

Összefoglalás

A Big Data lehetővé teszi az eddig nem ismert megoszlások vizsgálatát. Empirikus kérdés annak eldöntése, hogy különböző lehetséges mintázatok közül az adott halmazban melyiket találjuk. Tanulmányunkban két prosumer online közösség különböző zsánerekhez kapcsolódó aktivitási megoszlását vizsgáltuk. Bár voltak az egyes zsánerek között különbségek, domináns módon nem a Long Tail, hanem a Long Neck formáció dominált – vagyis az egyes zsánereken belül a könyvek 20%-a az aktivitások túlnyomó többsége. Miközben tehát a Long Tail kínálati oldala adva van, hiszen nagyszámú könyv elérését valószínűsíthetjük, addig a „keresleti” információ áramláshoz kapcsolódó hatást nem igazolja vissza az online közösségek tevékenysége. Ez nem csak azért érdekes, mert a szellemes és divatos fogalmak, mint amilyen a Long Tail is, végzete a túlterjeszkedés – amikor olyan kontextusokra is alkalmazzuk őket, amelyekben inadekvátak. Elméleti szinten mindez visszavezet a tanulmány elején idézett morenoi tétel igazságához: csakis a konkrét személyek konkrét hálózatainak konkrét viszonyinak megismerése után vállalkozhatunk az egyes csoport és társadalmi struktúrák működésének megértésére. Ez egyben egy új kutatási program megvalósításának szükségességét is kijelöli. Tanulmányunk ebben az irányba tett első lépésként is felfogható.

Irodalom

- Anderson, C. (2004): The Long Tail. *Wired Magazine*, October 2004.
- Anderson, C. (2006): *The Long Tail: Why the Future of Business is Selling Less of More*. New York. Hyperion
- Anderson, C. (2007): *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete*, *Wired* 2007. 16.
- Barabasi A. L. (2002): *Linked: The New Science of Networks*. Perseus Book Group.
- Bell, D. (1973) *The Coming of Post-Industrial Society: A Venture in Social Forecasting*, New York: Basic Books,
- Burt, R. S. (2000): “The network structure of social capital.” *Research in organizational behavior* 22 345-423.
- Brynjolfsson, E., Hu, Y., Simester, D. (2011): “Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales.” *Management Science* 57.8. 1373–1386.
- Castells, M. (1996): second edition, 2000). *The Rise of the Network Society, The Information Age: Economy, Society and Culture Vol. I*. Cambridge, MA; Oxford, UK: Blackwell.
- Castells, M. (1997): second edition, 2004). *The Power of Identity, The Information Age: Economy, Society and Culture Vol. II*. Cambridge, MA; Oxford, UK: Blackwell.
- Castells, M. (1998): second edition, 2000). *End of Millennium, The Information Age: Economy, Society and Culture Vol. III*. Cambridge, MA; Oxford, UK: Blackwell.
- Cowen, T. (2013): *Average is Over: Powering America Beyond the Age of the Great Stagnation*. Dutton Adult
- Csepeli Gy., Dessewffy, T. (2015): *Big Data. A technological change that will fulfil sociology*, *Review of Sociology*
- Dessewffy T., Láng L. (2015): Big Data és a társadalomtudományok véletlen találkozása a műtőasztalon, *Replika* 92-93. sz.
- Elberse, A. (2013): *Blockbusters: Hit-making, Risk-taking, and the Big Business of Entertainment* Henry Holt.

- Fowler, J. H., Christakis, N. (2010): *Kapcsolatok hálójában - Mire képesek a közösségi hálózatok, és hogy alakítják sorsunkat* Typotex Kiadó
- Jiang, B. (2015): Head/tail Breaks for Visualization of City Structure and Dynamics *Cities*, 43, 69-77.
- Kossinets, Gueorgi; Duncan J. Watts (2006): "Empirical analysis of an evolving social network." *Science* 311, no. 5757 88-90.
- Latour, Bruno (2010): Tarde's idea of quantification in: Candea, Matei ed. *The Social after Gabriel Tarde: Debates and Assessments*, Routledge
- Moreno, Jacob L. (1937): Sociometry in Relation to Other Social Sciences, *Sociometry*, Vol. 1, No. 1/2 (Jul. - Oct., 1937)
- Moreno, Jacob L (1947): Organization of the Social Atom *Sociometry*, Vol. 10, No. 3 (Aug., 1947)
- Orlikowski, W. J., Yates, J. (1994): "Genre repertoire: The structuring of communicative practices in organizations." *Administrative science quarterly* 541-574.
- Ritzer, G.; Jurgenson, N. (2010): Production, Consumption, Prosumption The nature of capitalism in the age of the digital 'prosumer'. *Journal of consumer culture*, 10(1), 13-36.
- Rogers, E. M. (2010): *Diffusion of innovations*. Simon and Schuster
- Toffler, A. (1980): *The third wave: The classic study of tomorrow*. New York, NY: Bantam.
- Yates, J., Orlikowski, W. (2002): "Genre systems: Structuring interaction through communicative norms." *Journal of Business Communication* 39, no. 1 13-35.
- Watts, D. J. (2004): *Six degrees: The science of a connected age*. WW Norton & Company.

