

A szociológia és a Big Data¹

Ha elolvassuk Huszár Tibor (2015) most megjelent, a magyar szociológia történetéről írott könyvét, elkap bennünket egy különös hangulat. Mintha nem az számított volna a múltban, hogy mit írtak, mit gondoltak a szociológusok, csak az számított, hogy *voltak*. Életüknek, műveiknek tétjük volt. Folyamatosan perben és haragban álltak a hatalommal, mely a maga szája íze szerint jutalmazta és büntette őket. A könyvben bemutatott szerzők közül többen rövidebb-hosszabb ideig börtönbe kerültek, voltak, akiket üldöztek, elhallgattattak, és akadtak olyanok, akik ideig-óráig maguk is a hatalom részesei voltak. A szociológia egyszerre volt a remény és a félelem forrása. Jól látható persze, hogy a hatalom kitüntetett, „double bind” érdeklődése nemcsak a szociológusok, hanem általában véve az orosz értelemben vett „intelligencia” szereplői iránt nyilvánult meg.

Mindez mára a múlté lett. Ma már a szociológia senkit sem érdekel, aki kívül van az oktatás, a kutatás, a konferenciák és szakmai publikációk világán. Tisztes iparrá lett a szociológia, mely művelői számára nyugaton tisztes, keleten kevésbé tisztes megélhetést biztosít. A mai szociológiában nincs semmi, amitől félni kellene vagy amiben reménykedni lehetne.

Rengeteg a kutatás, rengeteg a kutató, megállíthatatlan a publikációk árama. A sok eredmény azonban nem áll össze egységgé, nincs egy nagy elméleti narratíva, mely megfejtést kínálna a társadalmi lét talányaira. A szociológia és a társadalom elvált egymástól. A szociológiai kutatások eredményességét nem a társadalmi változások mutatják. Az eredményesség mutatója a publikációkra történő hivatkozások száma. A szociológia zárt diskurzushálózattá vált, melyen belül a skálafüggetlenség az új. Kevesen vannak, akikre sokan hivatkoznak, és sokan vannak, akikre alig vagy egyáltalán nem. Minden nemzedék újrakezdi a versenyt, nincsenek többé halhatatlanok. Az egykori halhatatlanok művei a szociológiatörténeti ipar tárgyai lettek, szellemük nem él, csak a nevük.

A szociológia képtelenné vált a változások nyomon követésére. Mire publikussá váltak az egyes társadalmi jelenségekre vonatkozó kutatások eredményei, addigra a helyzet gyöke-

1 A cikk a 'PIAC_13-1-2013-025 (Prediktív e-kereskedelmi megoldások kutatása online és offline tartalmak forgalmazása során keletkező nagy tömegű adattartalmak feldolgozását segítő fúziós módszertan fejlesztésére) pályázata keretében íródott.

resen megváltozott. Az eredmények semmire nem voltak jók, csak arra, hogy a változások felderítésére újabb kutatásokat lehessen lefolytatni, melyek sorsa ugyancsak a megkésetttség lett (Csepeli és Prazsák 2004).

A szociológiatörténet és a társadalomelmélet egyé vált. Nincsenek új, par excellence szociológiai elméletek, melyek a régié helyére léptek volna, miközben a régi elméletek elvesztették heurisztikus értéküket. Eltűnt a valóság, mely a régi elméleteket élte. Nincs többé „társadalom”, melynek volt „struktúrája”. A struktúra eltűnésével eltűnt az „egyenlőség” kék madara, melyet kergetve keseregni lehetett az „egyenlőtlenségeken”. A társadalmi valóság többé már nem kínálja a „rend” illúzióját, amire a szociológia korábban épített.

A változások legszembetűnőbb sajátossága, hogy cseppfolyóssá váltak a határok, melyeket alapul véve a szociológia korábban oly magabiztosan ki tudta jelölni tárgyait. A tárgyak helyén szavak által teremtett szemantikai terek maradtak, melyek határait a fuzzy logika jelöli ki, ami nem több, mint az adott szavakat használók hite abban, hogy amiről beszélnek, az van. Ki merné ma megmondani, hogy mi a „család”, a „nemzet”, az „etnikum”, a „bűn” és sok minden más, melyek vizsgálatára szociológusok ezrei vállalkoznak évről évre szerte a világban.

Hankiss Elemér utolsó, kéziratban maradt művében az emberi színjáték hangjaira és színeire süket és vak univerzumból ír, melyben a magára maradt emberiséget nem „védi a csilagos ég boltozata” (Hankiss 2014).

Az egymást követő szociológiai kutatási paradigmák mind belül maradtak a szociológiai diskurzuson, s egyik paradigma sem volt képes kilépni a „természet-társadalom” dichotómiából, melyen belül a szociológiának természetyszerűleg a „társadalom” jutott. Hankiss Elemér idézett tanulmányában ráérezett arra, hogy ez a dichotómia tovább már nem tartható. Ami az emberrel történik a létben, az minden más létezővel megtörténik, s nem a tudomány, hanem a művészet vagy a teológia dolga annak kiderítése, hogy mit kezd az ember a maga történeteivel, hol keresheti élete célját, értelmét, honnan meríthet reményt, s hol találhat magyarázatot szenvedéseire.

A szétparcellázódott szociológiai kutatás nem vett róla tudomást, hogy a 21. században kiteljesedett a globalizáció, ami nolens volens visszavonhatatlanul megváltoztatja minden szociológiai probléma kontextusát, tűnjön az bármennyire partikulárisnak vagy irrelevánsnak. A globalizáció kényszerítő lehetősége Marx óta aligha tagadható (Jensen 2014). Ugyanakkor az is igaz, hogy a globalizáció lehetőségéből valósággá csak a 21. századra vált. A globalizáció társadalmi valósággá válását paradox módon egy technológiai folyamat gyorsította fel, amelynek lényegét Castells ismerte fel. Castells három, egyenként kissé túlírt kötetben részletesen megfogalmazta, hogy az információ továbbítására alkalmas technológiák egy-egy platformra kerülése révén egy a Föld minden pontját elérő médium jött létre, mely magában egyesíti az összes korábbi kommunikációs architektúrát, s teszi ezt ráadásul visszamenőleg is (Castells 2005 [1996], 2006 [1997], 2006 [1998]).

Az internet, mert ez az új médium neve, egyedülállóan alkalmas közeg arra, hogy mindenki, aki rákapcsolódik, kommunikálhasson mindenki mással, aki ugyancsak részese a hálózatnak, miközben sem a tér, sem az idő nem korlátozza a kommunikációt közöttük. De ez csak az alap. Minden, ami az internet közegében történik, megmarad s visszakereshetővé válik. Az internet előtti korszakban keletkezett tudás is beköltözik az új médiumba. Az internet behatol a társadalom alrendszerébe. Az internet közegébe költözik az üzleti élet, a politika, a kormányzás, az oktatás, a gyógyítás. Az életvilág, mely korábban a magánélet

kitüntetetten intim, zárt köre volt, az interneten kinyílik a végtelenbe a megerősítésekre éhes én új terepeként.

Barabási László bizonyította be, hogy az internetre kapcsolódott emberek és az általuk létrehozott, befogadott tartalmak között kialakult kapcsolatok szerveződésének logikája kizárja az egyenlő esélyeket (Barabási 2003 [2002]). A hálózatban központok és perifériák vannak, melyek dinamikáját a már Merton által leírt, majd a hálózatkutatásban újra felfedezett Máté-effektus írja le: „akinek van, annak adatik, kinek nincs, attól elvételik” (Mt 13.10–17).

Az internet behatolása a társadalomba, az IKT globalizációs nyomása észrevétlen maradt a szociológusok számára, akik – mintha mi sem történt volna – továbbra is kutatták az egyre kisebb és kisebb körre szűkülő társadalmi problémákat, s hajsztolták az újabb és újabb „beszédmódokat”, diskurzusokat, melyek kizárólag áltál váltak elgondolhatóvá és megvitathatóvá, hogy a régi szavak helyére új szavak kerültek, de az új szavak jelentéstartománya is éppen úgy a fuzzy logika szerint alakult, mint a régieké.

A szociológia válsága természetesen a Big Data nélkül is tárgyalható, de itt és most azt a gondolatot szeretném felvetni, hogy a Big Data adta lehetőségek kihasználása hozzásegíthetné a szociológiát, hogy újra relevánssá váljék, ha nem is abban az életveszélyes értelemben, ahogy releváns volt Magyarország a 20. században.

Mi a Big Data? Mayer-Schönberger és Cukier jól látják, hogy a Big Data esetében a „big” jelző nem egyszerűen az adattömeg nagyságára, inkább korlátlan növekedési potenciáljára utal (Mayer-Schönberger és Cukier 2015). Az internetre került társadalmi folyamatok minden történése nyomot hagy maga után, ami adatként marad meg ott, ahova kerül. Az adatok folyamatosan bővülő adatbázisokba kerülnek, melyek külön-külön s egyesítve alapját képezhetik a Big Data alapú vizsgálódásoknak. Nyilvánvaló, hogy az adatok nyersen még nem alkalmasak a velük végzett munkára, különösen nem, ha strukturálatlanok, ha mint képek, szövegek, zenék kerülnek be az adatbázisba. A felhasználók, kliensek viselkedései viszont térben és időben strukturált adatokat hagynak maguk után. A kétféle adattípus összerendezése nem triviális feladat, de megoldható.

Látnunk kell, hogy az adattermelés nemcsak annak a következménye, hogy a társadalmi élet nagy része az interneten zajlik. A természetben zajló folyamatok nyomain is rögzíthetők, tárolhatók, elemezhetők, amennyiben megfelelő szenzorok és az általuk felvett adatok továbbítására alkalmas eszközök állnak rendelkezésre. Az egy helyen, egy időben zajló társadalmi és természeti folyamatok adatainak együttes elemzése a természeti és társadalmi valóság nem várt, teljesen új, gyakorlati és elméleti szempontból egyaránt lényeges következtetésekre alkalmas nyújtó modelljeinek megalkotására ad lehetőséget.

A Big Data-megközelítés szociológiai jelentősége a következőkben foglalható össze:

1. *Azonnalosság.* Az adatok felvétele és elemzése között drasztikusan lerövidül az idő, ami a gyakorlat szempontjából felbecsülhetetlen előnyt jelent. A szokványos szociológiai adatfelvételek esetében az adatok keletkezése és elemzése közötti idő hosszú, ami sok esetben a gyakorlat szempontjából eleve irrelevánssá teszi az eredményeket.

2. *Viselkedés.* Az adatbázisba olyan adatok kerülnek, amelyek közvetlenül az aktoroktól származnak, nincsenek kognitív szűrők, mint a kutatói hipotézis, a kérdőív, a kérdező és maga a kérdezett. Egy nagy telekommunikációs cég adattárházával megismerkedve még az ezredforduló idején alkalmam volt összevetni a valós idejű, valós telefonálásokra vonatkozó adatokat a telefonnal rendelkezők kérdőíves kikérdezéséből származó adatokkal. Az adattárházból vett adatok alapján készült elemzéseket pár nappal a keletkezésük után már

a vezetés elé tárhattuk, szemben a piackutatásból származó adatok alapján készült elemzésekkel, amelyek több hónapos késéssel kerültek a vezetők asztalára. De az igazi különbség az eredményekben volt. A kérdőíves kikérdezés eredményeként a telefonálók torzított képet adtak magukról, pontatlan vagy téves információkat adtak telefonálási szokásaikról, szemben az adattárházból származó adatokkal, melyek minden ügyfél esetében pontosan mutatták a hívások irányát, időtartamát, a hívottak és hívók közötti kapcsolatokat s persze a profit mértékét (Csepeli, Csere és Plisz 2002). A reprezentatív mintába került megkérdezettek a szociológus foglyai, hiszen csak arra válaszolhatnak, amiről kérdezik őket, és sokszor fogalmuk sincs, mit válaszoljanak, hiszen nincs hozzá szakértelmük (pl. honnan szerezzük be az atomreaktorba a fűtőelemeket) vagy éppen az a baj, hogy még nem gondolkodtak a számukra irreleváns kérdésekről.

3. *Teljeskörűség.* A szociológiai adatfelvétel minták kutatásán alapul, melyek a leggondosabb statisztikai feltételek esetében sem tudnak versenyezni a teljes körű adatfelvételekkel. A teljes körű adatfelvétel ugyanis lehetővé teszi egészen szűk szegmensek vizsgálatát is, amire a reprezentatív minták a kis elemszám miatt nem adnak lehetőséget.

4. *Longitudinalitás.* A társadalmi folyamatok időbeli alakulásának megbízható vizsgálata, ha ugyanazokat az aktorokat hosszú időn át nyomon követik. Ennek érdekében alkalmazzák a szociológiában a panelmódszert, mely azt jelenti, hogy egy adott minta tagjai hosszabb időn át a kutatók rendelkezésére állnak, s időről időre válaszolnak a kutatók által összeállított kérdésekre telefonon, személyesen vagy interneten. A panel tagjai azonban az idő multával egyre kevesebben lesznek, elköltöznek, megunják a kérdezést, meghalnak. A panelt persze a szociológusok kiegészítik, de az azonosság elve sérül ezáltal, hiszen nem lehet bizonyosan tudni, hogy az újonnan toborzott paneltagok mit válaszolhattak volna a kérdésekre akkor, amikor még nem kérdezték őket, mert nem tartoztak a panelbe. A Big Data esetében viszont semmiféle elvi nehézségbe nem ütköznek a korábbi viselkedésekre vonatkozó adatok lekérdezése és összevetése az azonnali adatokkal.

Megfelelő tisztítás, rendezés után az adatok származását jelölő változók szerinti megkülönböztetés rendjében a Big Data hatalmas és egyre növekvő adattömege kérdezhető aszerint, hogy mely változók mely változókkal mutatnak szignifikáns együtt járást. Mivel a változók száma eleve nem korlátozott (vagy csak az adatgyűjtő eszközök szabnak korlátot), elvben minden változó lehetséges együtt járása minden változóval vizsgálatra kerül, de a rostán csak a szignifikáns összefüggések maradnak fenn.

Egy másik elemzési irány, amikor csoportokat keresünk, melyek létezése a potenciális csoporttagok közötti hasonlóságokra és a távolságokra utaló adatok alapján mutatható ki. Fontos megjegyezni, hogy semmiféle előzetes megfontolás nem játszik közre a csoportok kialakításában, hiszen az adatok maguktól az aktoroktól származnak, s nem arra vonatkoznak, amit az aktorok magukról vallanak vagy amit a kutatók az aktorokról gondolnak, hanem azt mutatják, ami az aktorokkal ténylegesen történik (vásárolnak, telefonálnak, helyet változtatnak, vérnyomásuk, pulzusszámuk változik stb.). A szociológus által végzett klaszterelemzéshez képest az így kapott csoportok természetesebbek, mivel a priori szempontok nem játszanak szerepet a létrehozásukban. Ugyanakkor mindkét esetben a homofília-heterofília mentén jönnek létre a csoportok, s egyik esetben sem feltételezhetjük, hogy közösségként működnének.

Arra is lehetőség van, hogy a már létrejött csoportok szerint az újabb adatok elrendezhetők legyenek. Ez esetben a feladat az ismeretlen ismerőssé tétele. A Big Data a felismerést

segíti, aminek az elhárítással, nemzetbiztonsággal, bűnüldözéssel foglalkozó szervezetek veszik nagy hasznát.

Az egyes változók közötti együtt járások idősoros mintái különleges lehetőséget adnak arra, hogy becslésekkel lehessen élni a minták jövőbeli alakulásairól. Az előrejelzés természetesen sosem teljesen biztos, de a pusztán sejtéshez képest jelentős segítség a tervezéshez.

A szabályszerűségek megismerését követően a Big Data adattömegébe bekerült adatok elemzése a szabálytól való eltérésekre, a kivételekre, a szokatlan együtt járásokra is felhívhatja a figyelmet. Ez a funkció alkalmas a csalások, a kríziseket megelőző bizonytalanságok felderítésére, aminek különös aktualitást ad korunk, melyet nem alaptalanul a bizonytalanság korának nevezünk. A szabályszerűtől való eltérésben nemcsak a deviancia, de az újítás lehetősége is benne rejlik, amit Merton már említett nagy művében tökéletesen megérezett.

A Big Data korrelációs mátrixai aktuálisan hiányzó, de egyébként létező adatok kipótlására, becslésére is alkalmassá tehetők, s az újabb adatok érkezésével az eredmények ellenőrizhetők is lesznek.

Végül érdemes megemlíteni a szimulációt, melynek révén a meglévő változókkal számolva, különböző peremfeltételeket szabva az adattömegben reprezentált társadalmi és természeti folyamatok különböző lehetséges kimenetei megjeleníthetők lesznek. Ez a természetes kísérletezés előtt nyit új távlatokat.

Látszólag formai, ugyanakkor a felhasználás szempontjából lényegi jellemzője a Big Data alapú elemzéseknek az eredmények vizualizációs megjelenítése, amit a szöveges megjegyzések értelmeznek, kiegészítenek. A jól sikerült vizualizáció nem csak interpretációs, hanem analitikai eszköz is egyben. A következtetéseket azonban a felhasználónak magának kell levonnia.

A Big Data annak arányában jut majd uralkodó szerephez a társadalmi és természeti folyamatok, valamint a köztük létesülő interakciók megismerésekor ahogyan egyre inkább kiterjed a kör, ahol adatok keletkeznek, továbbításra kerülnek s egységes platformokba rendeződnek. A végeredmény a természeti és társadalmi folyamatok korábban elképzelhetetlen kiterjedésű és komplexitású modellezése, ami nem elsősorban az Egész, hanem azon belül részek, sőt egészen kis részmozzanatok, köztük az egyes egyének szempontjából vezet sorsdöntő következményekhez. Gondoljunk csak arra, hogy a mindennapokban élő ember teste kulcsfontosságú szerveinek – köztük a szív és az agy működési adatainak – monitorozása, központba továbbítása milyen lehetőségeket kínál a betegségek korai előrejelzésére és a gyógyító beavatkozások időben való elkezdésére. Az üzleti életben a Big Data alapú elemzések, ajánló rendszerek a piac legapróbb rezdüléseire érzékeny kínálati struktúra kialakítását teszi lehetővé, kiélezve az ügyfelek elégedettségéért folytatott versenyt, mely nemcsak a jelen, hanem a közeljövő ügyfeleiért is folyik. A politikában a Big Data lehetővé teszi a választók közvetlen és közvetett politikai aktivitásának nyomon követését, modellezését, az üzenetek élesben történő tesztelését, a „nanocélzást”, ami egészen szűk, hálózati szempontból kulcsfontosságú szavazói rétegek elérését teszi lehetővé. De lehetőség nyílik a hatalom kontrolljára is (Wikileaks). A társadalmi média s annak zászlóshajója, a Facebook már ma is a Big Data alapú marketing terepe. A „dolgok internete” tovább bővíti a megismerés körébe vont valóság terjedelmét.

Mindennek alapján hiba volna arra a következtetésre jutni, hogy a szociológia és a szociológusok ideje lejárt. A Big Data megjelenése óriási kihívás a szociológusok számára, akiknek bele kell nyugodniuk, hogy korábbi kedvenc szerzőik és korábban alkalmazott kedvenc ku-

tatási módszereik történelemmé válnak, ugyanakkor észre kell venniük, hogy az algoritmusok nem tudnak gondolkodni. A korrelációkba, a klaszterekbe, a szabályszerűségektől eltérő mintázatokba, a forgatókönyvekbe más, mint a szociológus, nem tud értelmet lehelni.

Az új társadalomelmélet, mely a Big Data eredményeit a társadalmi folyamatok megértésének és magyarázatának építőköveivé teszi, nem lehet normatív s nem terjeszkedhet túl illetékességi körén, mivel minden egyes állítása azonnal bizonyítható vagy cáfolható. Az új társadalomelmélet voltaképpen egy tanuló algoritmus, mely a szociológusok kollaborációjában alakul és fejlődik, kiszzelektálva az aktuálisan érvényes és az érvénytelen állításokat. Ez az elmélet – az idealisták legnagyobb fájdmájára – mit sem tud kezdeni a társadalmi valóságot megkettőző felfoggással, mely szerint a lét és a tudat, az objektivitás és a szubjektivitás struktúrái örök harcban állnak egymással. A Big Data tükrében csak egyetlen, a preszokratikus filozófusok által már megidézett valóság van, melyben megszakítatlanul zajlanak a történések az emberekben, az emberek között s az embereken kívül. A Big Data-ra alapozott valóságfelfogás ugyan elveszíti a réven a szubjektivitást és a tudatot, de a gyakorlatban a vámon visszadja azt, hiszen olyan tudást tesz elérhetővé az üzleti és politikai döntéshozók, hálózatihub-emberek számára, aminek alkalmazása a mindennapokban élő embereket egészségesebbé, elégedettebbé, sőt talán boldogabbá teheti.

A tudás, melyről beszélünk, ugyanakkor nemcsak a béke, hanem a háború eszköze is. Fantasztikus, korábban elképzelhetetlen fegyvereket ad a világ színpadán a javak újraelosztásáért, a befolyásért, a globális hataloméért küzdő erők kezébe. A Big Data önmagában véve sem nem jó, sem nem rossz, léte azonban növeli a Jó és a Rossz közötti örök harc tétjeit.

Hivatkozott irodalom

- Barabási Albert László (2003 [2002]): *Behálózva. A hálózatok új tudománya*. Budapest: Magyar Könyvklub.
- Castells, Manuel (2005 [1996]): *A hálózati társadalom kialakulása. Az információ kora*. I. kötet. Budapest: Gondolat.
- Castells, Manuel (2006 [1997]): *Az identitás hatalma. Az információs kor. Gazdaság, társadalom, kultúra*. II. kötet. Budapest: Gondolat.
- Castells, Manuel (2006 [1998]): *Az évezred vége. Az információ kora. Gazdaság, társadalom, kultúra*. III. Budapest: Gondolat.
- Csepeli György, Csere Gábor és Plisz Judit (2002): A „legjobb csomag modell”. Piaci döntések szimulációja valóságos viselkedésre vonatkozó adatok alapján. *Magyar Távközlés* (2): 22–25.
- Csepeli György és Praszák Gergő (2004): Paradigm Change in Sociology. *Review of Sociology* 10(2): 39–46. Interneten: http://www.szociologia.hu/dynamic/RevSoc_2004_2_CsepeliGy_Paradigm_change.pdf
- Csepeli György és Dessewffy Tibor (2015): Big Data mint a szociológiai megismerés új paradigmája. *Szellem és Tudomány* (előkészületben).
- Csepeli György és Dessewffy Tibor (2015): Big Data. A Technological Change that Will Fulfil Sociology. *Review of Sociology* (nyomdában).
- Hankiss Elemér (2014): *Kvantummechanika és az élet értelme*. (Kézirat.)
- Huszár Tibor (2015): *A magyar szociológia története*. Budapest: Osiris.
- Jensen, Jody (2014): *Globalizáció és új kormányzás*. Budapest – Érd: MTA TTKK – Mundus-Novus.
- Mayer-Schönberger, Viktor és Kenneth Cukier (2015): *Big Data. Forradalmi módszer, mely megváltoztatja munkánkat, gondolkodásunkat és egész életünket*. Budapest: HVG.
- Merton, Robert K. (1980): *Társadalomelmélet és társadalmi struktúra*. Budapest: Gondolat.