

DR. HABIL. JAKOBI ÁKOS PhD

ADJUNKTUS

EÖTVÖS LORÁND TUDOMÁNYEGYETEM TERMÉSZETTUDOMÁNYI KAR

FÖLDRAJZTUDOMÁNYI KÖZPONT REGIONÁLIS TUDOMÁNYI TANSZÉK



Big spatial data, avagy új lehetőségek a területi döntés-előkészítés támogatásában

1. Bevezetés

A területi igazgatásszervezés, a területi tervezés, a területfejlesztés, a vidékfejlesztés vagy bármely más területi szempontból érzékeny szakpolitika ilyen vagy olyan szintű döntéseinek hátterében mindig valamilyen a térbeli szempontokat is figyelembe vevő szakmai előkészítő munka áll. A döntések nem spontán jellegűek, hanem körültekintő elemzésekre épülnek. A döntés-előkészítő munka javarészt statisztikai, felmérési adatokra, információforrásokra stb. támaszkodik, ám napjainkban – úgy tűnik – újabb adatforrások, és ezáltal újszerű szempontok figyelembevételére is lehetőség nyílik, ami a köznyelvben egyre inkább terjedő big data jelenséggel kapcsolható. A közigazgatás persze részben már felfedezte magának a big data jellegű forrásokban rejlő új lehetőségeket, gondoljunk csak a különböző szervezeteknél közfeladataik ellátása során keletkező adatok kezeléséről és feldolgozásáról szóló Nemzeti Adatpolitikára, illetve az azt megfogalmazó ún. Fehér Könyvre (NHIT 2016), ami viszont csak részben taglalja a kérdéskör területi dimenziókat érintő szempontjait. A térbeli tartalommal is rendelkező big data állományokban ugyanakkor hatalmas potenciál rejlik, melyekről érdemes néhány példa kíséretében több szót is ejteni.

Általánosságban fogalmazva a szakmai körökben „big data” (szabad fordításban „óriási adathalmaz”) néven ismert kifejezés arra a hatalmas adatmennyiségre utal, amely információs világunkban rapid módon és folyamatosan keletkezik, s melynek feldolgozása a hagyományos kapacitásokkal és módszerekkel már-már megoldhatatlan kihívást jelent. A big data ezzel ellentétben mégis nagy lehetőségeket kínál. A sokáig csak virtuális melléktermékként számon tartott napi információhalom ugyanis épp akkor válik értékké, amikor a különböző adatokat sikerül összekötni, köztük összefüggéseket, felismerhető mintázatokat találni, s mindebből értékelhető következtetéseket levonni. A kormányzati szerveknél, internetes és telekommunikációs cégeknél összegyűlő hatalmas adatmennyiség a társadalom és a gazdaság folyamatainak vizsgálata számára valóságos aranybánya. Közlekedési, vásárlási, szabadidős vagy épp mindennapi szokásainkról

szinte korlátlanul gyűjtenek adatokat a különféle szervezetek, s köztük a közfeladatokat ellátó intézmények is igen nagy számban. Mindezek összessége végső soron betekintést nyújt az emberi viselkedés egyedi és csoportos (társadalmi) szintjeibe is. Az egyre szélesebb körben terjedő, térbeli információkat is használó vagy azokat spontán mód generáló alkalmazások, digitális eszközök és webes szolgáltatások révén a big data a társadalom térbeli működésének megértéséhez vagy feltérképezéséhez is megszámlálhatatlan mennyiségben kínál új adatokat, s végső soron ezek az információk a területi tervezési vagy igazgatási döntések megalapozását segíthetik.

A technológiai fejlődés, az új IKT innovációk megjelenése, továbbá a digitális eszközök egyre szélesebb körű megfizethetősége mind-mind hozzájárultak ahhoz, hogy beindulhasson a big data jelenség néven ismertté vált adatrobbanás akár a mennyiség, akár a diverzitás vonatkozásában. Egyes megfogalmazásokban ez nem más, mint az adatok ipari forradalma (Hellerstein 2008), melynek hatásaként az emberi viselkedés megértésének eddig soha nem látott történelmi korszakába léphetünk (Onnela 2011). Ráadásul ezen új adatok és adatforrások megszámlálhatatlan új lehetőséget is magukkal hoztak a társadalmi, gazdasági vagy egyéb direkt szakpolitikai döntések támogatásához. Persze a megszámlálhatatlan mennyiségű adat érdemleges információhalmokba rendezése és a trendek, mintázatok kiszűrése vagy meghatározása komoly számítási kapacitásokat igényel, melyek csak újabban kezdenek rendelkezésünkre állni. A kinyert eredmények ugyanakkor új megvilágításba helyezhetik az eddigi hivatalos adatgyűjtési forrásainkat, sőt egyre komolyabban kiegészíthetik azokat az emberi viselkedés kutatásának mélyebb vagy részletekbe menő motívumaival.

Maga a big data terminus még nem teljesen tisztázott, bár összefoglaló szakmai kísérletekről már olvashatunk (Szűts – Yoo 2016). Egyesek pusztán a hatalmas adatfájlokat értik alatta, ami főként a társadalomtudósok gyakori véleménye. Az informatikusok azonban másként vélekednek: ők az adattáramlásra és a folyamat-generált adathalmazokra gondolnak elsősorban. Sarkosan fogalmazva, míg a korábbi időszakban csak relatíve kis mennyiségű analóg adat keletkezett és vált

elérhetővé a korlátozott számú csatornán keresztül, addig ma rendszeresen nagyon nagy mennyiségű digitális adat generálódik számos különböző csatornán keresztül minden egyes percben (King 2011). Egyrészt az adatok keletkezésének és átvitelének sebessége és gyakorisága, másrészt a források száma és sokszínűsége az, ami az adatok új özvívyszerű áramlását jellemzi. Erről számos szakmai anyagban olvashatunk részletesebben (pl. Giczi – Szőke 2017).

Napjaink rohamos információtechnológiai fejlődése olyan szolgáltatásokat és alkalmazásokat is életre hívott, amelyek a területi kutatás, a területi igazgatás és a területi tervezés számára is merőben új eszközöket és lehetőségeket kínálnak. Az új lehetőségek leginkább abból fakadnak, hogy az információs és kommunikációs technológiák és a számítógépes megoldások már-már az élet szinte minden szféráját áthatották. Az angol nyelvű szakirodalomban „*pervasive computing*” vagy „*ubiquitous computing*” terminusokkal jelölt fogalom (Satyanarayanan 2001, Friedewald – Raabe 2011, Weiser M. 1991) háttérben pedig a térbeli információs eszközhasználat szétterjedése és a térbeli információrobanás is megfigyelhető (Jiang – Yao 2006, Galloway 2004, Zook et al. 2004). A területi kutatók és tervezők számára éppen abban rejlik új potenciál, hogy ezek a térbeli információk egyre szélesebb körben válnak hozzáférhetővé és a társadalom mind sokszínűbb térbeli működési mechanizmusába kínálnak bepillantást vagy legalábbis jobb rálátást. Sőt, ezen új adatok nem is pusztán elérhetőek, de úgyszólván rá is zúdulnak manapság a területi kérdésekkel foglalkozó szakmára. A világhálón közzétett strukturált vagy struktúrálatlan térbeli információtartalom vagy az egyre szélesebb körben terjedő, térbeli információkat is használó alkalmazások révén a big data a társadalom térbeli működésének megértéséhez is megszámlálhatatlan mennyiségben kínál új forrásokat. Terjedőben van tehát a big data fogalomkörön belül a külön térbeli tartalommal bíró adatok „*big spatial data*” elnevezése.

Amiket a big spatial data környezetben a területi kutatói és tervezői kör elsősorban haszonnal vizsgálhat, azok az úton-útfélen hagyott térbeli tartalommal is rendelkező direkt vagy indirekt digitális nyomok. Az efféle adatokra épülő adatbázisok közvetlen módon például az okostelefonok különböző helyalkalmazásaihoz kötődően keletkeznek (ezeket végső soron a felhasználó állíthatja be), vagy például egyes honlapok célzott geotag-ekkel, azaz földrajzi azonosító kódokkal való ellátásakor. Ám ennél is jóval érdekesebbek a geoinformációkat tartalmazó digitális nyomok indirekt halmazai, melyek nem szándékoltan, de mégis nagy számban keletkeznek. Példaként említhetők azok az elektronikus közlekedési kártyák vagy megfigyelő rendszerek, amelyek rögzítik a közlekedési rendszerbe való belépés és kilépés helyét és idejét, lehetőséget adva – elméletileg – a közlekedési térpályák, szokások stb. vizsgálatára. Digitális nyomokat hagyunk továbbá akkor is, amikor egy-egy weboldalt meglátogatunk, hiszen (általában) beazonosítható az az IP-cím, s ezáltal az a földrajzi hely is, ahonnan a világháló szolgáltatásait igénybe vettük. A digitális nyomok indirekt felhasználására, elemzésére persze számos más példa is említhető (például Girardin et al. 2009, Järv 2012, Naaman 2011), melyek mind a „*mel-*

léktermékként” keletkező digitális adatok vizsgálatával hozták meg következtetéseiket.

Folyamatosan változó világunkban az információtechnológiai eszközök és alkalmazások piacán sorra jelennek meg az újabb és újabb termékek és ezzel egyidejűleg folyamatosan módosulnak a felhasználói igények és szokások is. Ezen változások egyik nem elhanyagolható következménye, hogy a területi információk iránti igények is jelentős mértékben növekednek, és egyre több olyan eszköz és alkalmazás lát napvilágot, amely közvetlen vagy közvetett módon épp ezen területi információk hasznosítója vagy előállítója. Mindez legfőképp (de nem csak) a helymeghatározó eszközökkel egybeépített okostelefonok rohamos terjedésének köszönhető, melyek révén a geoadatok használatát akár már a mindennapok részének is lehet tekinteni. Az új infokommunikációs eszközök egy jelentős része nemcsak információtovábbításra, de kifejezetten területi adatok gyűjtésére is alkalmas, sőt a felhasználók egyre növekvő hányada egyenesen keresi a térbeli információkat is kínáló alkalmazásokat. Az okostelefonok azok számára is mindennapossá tették a térérzékeny adatok használatát, illetve annak lehetőségét, akik korábban nem sok ambíciót éreztek eziránt. Ezek a készülékek általában beépített helymeghatározó alkalmazásaikon keresztül aktív téradat-használati lehetőséget kínálnak, közelebb hozva ezzel a GPS technológiák nyújtotta lehetőségeket a felhasználókhoz. A telekommunikációs fejlődési trendekhez az is hozzátartozik, hogy az elkülönült céleszközök (GPS készülékek) piaca egyre szűkül, míg a mobil szoftveres piac bővül. Az Ipsos Zrt. 2010-es felmérése szerint Magyarországon a mobilkészülékek 22%-a rendelkezett már GPS készülékkel, s a felhasználók több mint fele már akkor elképzelhetőnek tartotta, hogy a mobil készülékét használni fogja három éven belül helymeghatározásra. A felmérés a lokalitást használó appok alkalmazásának piaci potenciálját is kimutatta, ugyanis a megkérdezett felhasználók kb. fele kisebb vagy nagyobb összeget fizetett volna is azért, hogy efféle szolgáltatást igénybe vehessen (Pintér 2011). Az azóta eltelt rövid időszakban ezek a jóslatok nagyrészt igazolódtak.

Az okostelefonok és a hasonló eszközök és alkalmazások révén azonban nemcsak a térbeli adatokhoz való hozzáférés lehetősége változott meg. Az alkalmazói kör akarva vagy akaratlanul milliányi digitális nyomot hagy maga után nap mint nap, melyek jelentős része épp földrajzi tartalmú, és melyek mellett a területi kutatói és tervezői szakma sem mehet el szórtanul.

A big data korszak új adatforrásai, illetve adatgyűjtési formái áttételesen többek között az információs kor területi differenciáltságáról is tájékoztatást adhatnak, ezek pedig a területi döntés-előkészítés alapjául szolgálhatnak. Már az is, hogy hol, milyen térségekben keletkezik ez a hatalmas adatmennyiség, informálhat minket az egyes területek virtuális világbeli jelenlétének abszolút súlyáról. A big data felhőjében kirajzolódó digitális lábnyomok térségenkénti nagysága, minőségi eltérései pedig az információhasználó lakosság területi jellemzőiről szolgálhatnak új adalékokkal vagy a legkülönbözőbb térségi szintek folyamatainak megértéséhez adhatnak támpontokat.

2. Néhány új lehetőség a területi döntés-előkészítés támogatásában

A big spatial data állományokra épülő analitikai lehetőségek köre szinte végtelen. Köztük számos olyan eljárást vagy módszert találhatunk, amelyek kifejezetten jól segíthetik a területi döntés-előkészítő munkát. Az alábbi példák csak apró szeletét mutatják be azoknak a lehetőségeknek, amelyeket a térhasználati információkra épülő várostervezés vagy a területi tervezés felhasználhat.

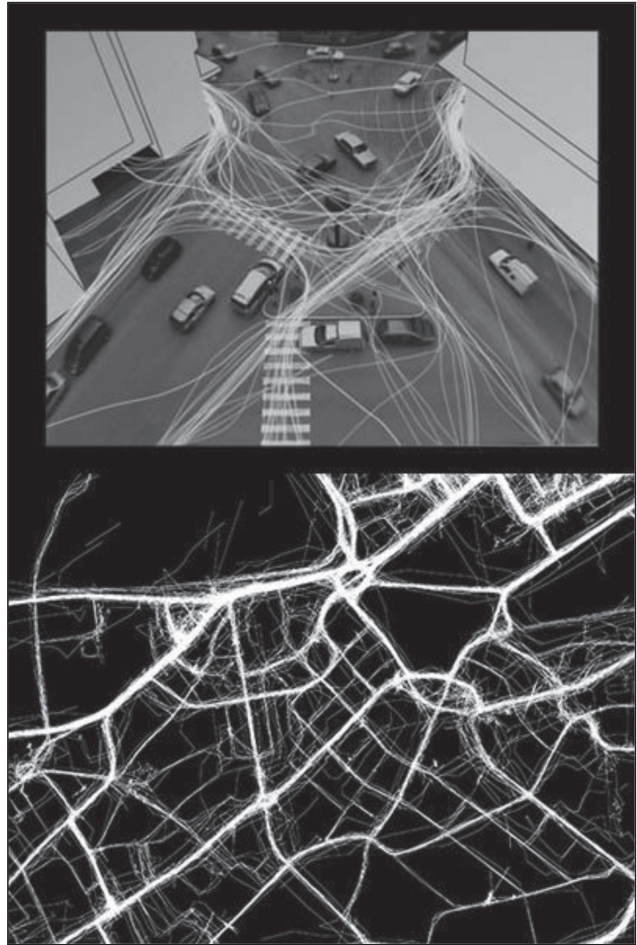
A digitális nyomokban rejlő sokszínű térbeli információk egyik széles körben elterjedt praktikus felhasználói területét képezik azok a megoldások, amelyek a helymeghatározó vagy egyéb rendszerekben rögzített direkt helyzeti adatok idősoros feldolgozását állítják a fókuszpontba. A nagyszámú tér-idő adatot hasznosító vizsgálatok tipikus eredményeiként a felmért egyedek vagy csoportok jellemző térbeli mozgási mintázatait, azaz térhasználati szokásait és tipikus bejárt térpályáit is meghatározhatók lehetnek.

Általánosságban a helymeghatározó, illetve navigációs eszközök segítségével a térhasználati döntéseket nemcsak befolyásolhatjuk, de értékelhetjük is, s az adatgyűjtési technológiákkal kiegészítve a térpályák jellemzőit jobban meg is ismerhetjük. A modern IKT eszközök immáron bőséges új forrásokat kínálnak a térhasználati szokások feltérképezéséhez, valamint a térpályák objektív felméréséhez. A GPS eszközök, az okostelefonokba beépített alkalmazások, illetve számos más helymeghatározó, nyomkövető (tracking) stb. alkalmazás képes rögzíteni az egyes felhasználók által megtegyt utvonakat, illetve azok tér-idő koordinátáit.

A geokoordináták és időbélyegek sorozataként rögzített tér-idő adatok különösen akkor válnak hasznossá, amikor nagyon sok felhasználó nagyon nagy számú hasonló információját sikerül egyszerre, áttekinthető módon értékelni. A big spatial data rendszerek éppen ezt kínálják. A sokfelhasználós közegekben ugyanis nemcsak az egyedi térpályák, hanem a jellemző utvonak, térbeli csoportosulások meghatározására is lehetőség nyílik és ezt már a várostervezési és településfejlesztési döntés-előkészítő folyamatokban is hasznosítani lehet.

Amióta a helymeghatározó rendszerek megoldásai széles körben és relatíve olcsón elérhetővé váltak, egyre nagyobb mennyiségű felhasználóról váltak ismertté a naplózott helyzeti információk, melyek az adatfeldolgozás után térképi formában is megjeleníthetők és elemezhetők (1. ábra). Ez a módszer (a térbeli azonosítás pontosságától függően) így lehetővé teszi például a városi mikro- vagy makroterek jellemzően vagy intenzíven használt helyeinek és utvonainak feltárását, kvázi mérését. A big spatial data forrásokra alapozott térpályá-analízis a településfejlesztés, várostervezés vagy közlekedéstervezés hasznos eszköze lehet, hiszen így már nem csak egyedi és alkalmi, de nagyszámú és szinte folyamatos megfigyelése is lehetséges az egyes helyeknek, tereknek vagy utvonaknak. Az adatvolumen pedig következképpen a vizsgálatok robusztusságának fokát növelheti.

A városi térhasználat lenyomataként értelmezhető direkt helyzeti adatok mellett a társadalmi térbeli viselkedés megértésében igazán újszerű és egyben jelentős potenciált kínálhat



1. ábra

Nyomkövető alkalmazások által rögzített térpályák a városi térben, a várostervezés vagy közlekedéstervezés szolgáltatásban

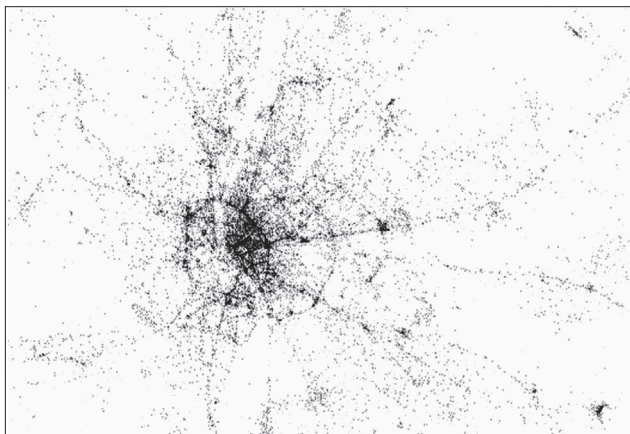
Forrás: Carden – Coast (2005)

az indirekt helyadatokat, melyek kihasználása az igazgatásszervezésben és a területi döntések előkészítésében még meglehetősen gyerekcipőben jár. Manapság egyre több olyan internetes felület létezik, ahol a közzétett információk egyértelműen azonosított földrajzi lokalizációval is rendelkeznek, aminek hatására az amúgy a tértől teljesen független információk is térbeli relevanciájuk lettek. Ha a HTML webkörnyezetben rögzült földrajzi lokalizációs szándékkal létrehozott geotag-ek (földrajzi azonosító kódok) adatait összegyűjtjük és térképen ábrázoljuk, láthatóvá válnak azok a területek, amelyekről nagyobb sűrűséggel tesznek közzé a felhasználók térbeli információkat vagy ahol gyakrabban megfordulnak, s így közvetve az is megvizsgálható, hogy hol használják aktívakban az információtechnológiai eszközöket, vagy melyek azok a térbeli pontok (helyek), amelyek a térszervezés szempontjából kitüntetett jelentőségűek.

A fellelhető számos példa közül csak egyet említve az utóbbi évek geolokalizált webes tartalomanalitikai kutatásainak népszerű forrásaként említhető a Twitter közösségi oldala, mely bepillantást nyújt a valóságos és spontán térszerveződések települési szintjeibe is. A közzétett rövid üzenetek (tweet-ek) rendkívül nagyszámú, bár nem teljes mintája egy nyilvános API (alkalmazásprogramozási felület vagy alkalmazásprogramozási interfész, angolul application program-

ming interface) segítségével adatbázisokba gyűjthető és szerkeszthető. A földrajzi azonosítóval is rendelkező üzeneteknek pedig a térbeli sűrűsödése, de ezen kívül egyéb tartalmi jellemzőinek területi előfordulási gyakorisága is elemezhető.

Mivel a Twitter üzenetek rendkívül precíz, akár az utca-hátszám szintnél is pontosabb területi azonosító geotag-ekkel rendelkeznek, esetenként lehetőség nyílik a pontsűrűségi adatok igazán részletes vizsgálatára. Gyakori változat, amikor a sűrűségi ábrák a területi azonosításra szolgáló háttérinformációk (közigazgatási beosztások, felszínrajz stb.) kirajzolása nélkül jelenítik meg a geokódolt találati pontokat. Az ábrákról ugyanakkor így is sejthető, hogy azok mely földrajzi térségekre vonatkoznak, mintegy közvetett módon visszaadva a települési vagy közigazgatási térfelosztás amúgy is meglévő releváns határait vagy objektumait. Ilyenkor a pontmintázat spontán kirajzolódásának érzékeltetése lehet a területi analitika célja, azaz például annak bemutatása, hogy a weben közölt tartalmak mennyire is követik a valódi földrajzi megkötöttségeket.



2. ábra

Geotagelt Twitter bejegyzések sűrűsége Budapesten
Forrás: Fischer (2013) alapján saját szerkesztés

A 2. ábra a Twitter bejegyzések Budapest környéki sűrűsödését jeleníti meg. Az ábrán dominánsabb, sötétebb színnel jelennek meg azok a helyek, ahol nagyobb gyakorisággal tettek közzé információkat a felhasználók. A térképen különösebb jelmagyarázat, vagy segédfelirat nélkül is kivehető a pesti belváros, a nagykörút, vagy egyes főbb utak sziluettje, de például nagyobb számban lehetett bejegyzéseket regisztrálni a Liszt Ferenc repülőtér környékén is. Fontos megjegyezni, hogy a kirajzolódó mintázatok nem feltétlenül a felhasználók számának egyenlőtlenségeit ábrázolják, inkább a térbeli tartalommal rendelkező információk relatív sűrűsödéseit, azaz a térségek tartalmi súlyát a virtuális térben. Persze ezek hátterében az abszolút volumenek, tehát az, hogy hol fordul elő általában több ember, meghatározók lehetnek, továbbá az is, hogy hol találhatnak a felhasználók említésre, közlésre érdemes mondanivalót, mely utóbbi leginkább az idegenforgalmi látványosságok közelében lehet felülreprezentált. A települési fejlesztések (például lokális idegenforgalmi fejlesztések) döntés-előkészítő lehetőségei mellett az eszköz így a települési tervezés egyfajta utólagos monitoring eszköze is lehet.

3. A hálózati térszerveződés értékelésének tapasztalatai

A big data források napjaink egyik legdinamikusabban fejlődő szeletét képviselik az online közösségi hálókban nyert adatok. Az efféle adatok felhasználási lehetőségei nemcsak a médiakutatók számára lehetnek érdekesek, de a társadalomkutatás számos más területe számára is, sőt – mint azt az alábbi példák is bizonyítják – a területi igazgatás és területi tervezés is hasznos információkhoz juthat ezen adatok felhasználásával. A fókuszban itt ugyanis a virtuális hálózaton keresztül realizálódó térszerveződés áll, pontosabban a mindennapi térbeli kapcsolatok kibertérbeli realizációi, avagy lenyomatai, melyek megismerése a külső térbeli kapcsolati formációk megértésében lehet kiemelten fontos (ezekre pedig épp a térbeli döntéseink alapulnak). Az online közösségeket vizsgáló korábbi tanulmányok ugyanis azt hangsúlyozták, hogy a virtuális világ–fizikai világ összekötöttséghez (Benedikt 1991) hasonlóan az internetes közösségek is térbeli, valamint társadalmi megkötöttségűek (Jones 1995). Az online közösségeket kulturális, strukturális, politikai és gazdasági jellemzők konstruálják, melyek földrajzi vonatkozásokkal átszőtt társadalmi kapcsolatokon és intézményeken alapulnak (Fernback 2007). A kapcsolati hálózatok ilyen értelemben tehát a spontán közösségi térszerveződés lenyomatai, melyek mintegy visszaigazolásul vagy épp cáfolatul összevethetők akár a közigazgatási térszerveződés mai rendszereivel.

Az online közösségi hálózatok (online social networks: OSN) lényegében virtuális kreálmányok, melyek a való élet baráti kapcsolatainak virtuális térbeli kifejezésére szolgálnak. Manapság mindemellett az IKT-alapú kommunikáció egyik jelentős új platformjának is tekinthetők, lehetővé téve a helytől független társadalmi életet is, habár a legújabb kutatási eredmények mégis arra utalnak, hogy a felhasználók és online barátaik való világbeli földrajzi helyzete mégis meghatározó lehet a hálózati topológia formálódásában. (Takhiteyev et al. 2012). Az online közösségi hálók így vélhetően egyidejűleg tükrözik vissza a kibertér és az offline földrajz társadalmi-területi jelenségeit. A közösségi hálók földrajzi helyekhez kapcsolása voltaképpen már azáltal megtehető, hogy maguk a felhasználók is lényegében földrajzilag azonosíthatók. Ha sikerül tehát kinyerni a geolokációs adatokat egy OSN hálózati adatbázisából, akkor lehetőség nyílik a felhasználók földrajzi kapcsolatrendszerének elemzésére és megjelenítésére is.

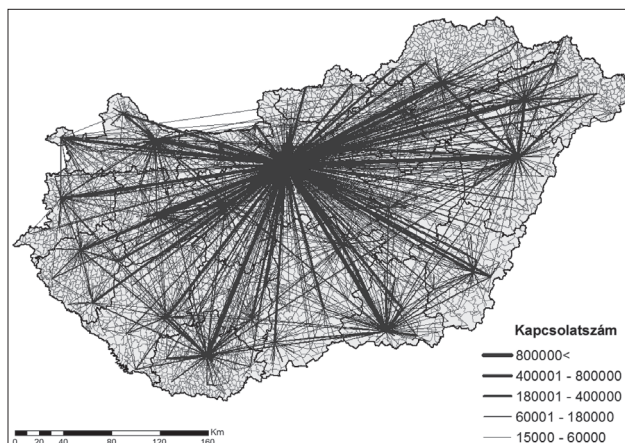
Az online közösségi hálók olyan nagyméretű hálózatok, amelyekben a felhasználók képezik a csomópontokat, élekek pedig a közöttük lévő kapcsolatok tekinthetők. A közösségi hálók web-alapú szolgáltatások, amelyek lehetővé teszik, hogy a felhasználók láthatóvá tegyék társadalmi kapcsolatrendszerüket (Boyd – Ellison 2007). A közösségi hálók az emberek közti kommunikáció kiegészítő formái, mégpedig főként azok között, akik már korábban, a valódi világban is ismerték egymást (Ellison et al. 2006). Másképp fogalmazva, a közösségi oldalak bár arra is használhatók, hogy új személyeket ismerjünk meg, azonban döntő többségben azokkal az emberekkel való kapcsolatainkat dokumentáljuk rajtuk, azokkal az eddigi offline világ hálózataiban kapcsolatba kerültünk. Ezek a kapcsolati hálók viszont, úgy tűnik,

térbeli megkötöttségeket is mutatnak. A területi tervezés vagy igazgatásszervezés döntés-előkészítő munkájában pedig az OSN-ek épp ezen tulajdonságát lehet kihasználni: a közösségi hálók az emberek többé-kevésbé valós társadalmi kapcsolatainak nagymintás lenyomatai, sőt áttételesen térbeli lenyomatai is, melyekről egyébként meglehetősen nehéz lenne más módszerekkel információkat szerezni.

A fentiekben említett elgondolások igazolásaként az egykor volt legnagyobb magyarországi online közösségi oldal, az iWiW adatait érdemes görcső alá venni. A következőkben bemutatásra kerülő vizsgálatok adatbázisában a felhasználók és a kapcsolatok száma lényegében minden településre hozzáférhető volt. A felhasználók földrajzi lokalizálásához a regisztrált személyek profil-információit lehetett felhasználni, melyek ugyanakkor alkalmanként problematikusak is lehetnek a lokalizációs kísérletekben (Hecht et al. 2011). Az iWiW-en ugyanakkor kötelező lokalitást megjelölni, amit a felhasználó egy választható listában jelölhet meg. A lakhely ezen adata később könnyen módosítható és természetesen nem kapcsolódik hozzá semmiféle valódi ellenőrzés. Ennél fogva joggal gondolható, hogy a felhasználók lokalizált adatai esetünkben néha bizonyos vonatkozásokban torzítottak lehetnek. Ugyanakkor az adatok nagy száma miatt a hibás rekordok aránya relatíve kicsinek becsülhető, így a társadalmi térszerveződés vizsgálatára egyértelműen alkalmas lehet ez az adatbázis. Az online közösségi háló területi sajátosságainak vizsgálatára kialakított adatbázisban összesen 4 058 505 felhasználó adata szerepelt, akik 2562 településen voltak regisztrálva. A felhasználók kapcsolati hálója összesen majdnem 786 millió összeköttetést tartalmazott, melyek közül kb. 370 millió a településhatárokon belül maradt, míg kb. 416 millió két különböző település lakója között jött létre. A hálózati adatbázis adatait összesen 1,37 millió település-település kapcsolat (él) formájában lehetett aggregálni.

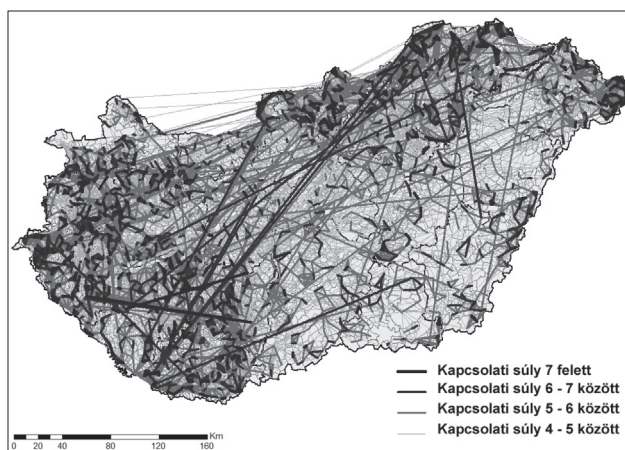
A fent említett adatbázisra építve teszteltük a területi igazgatás és a spontán térszerveződés kapcsolatát a településhierarchia kirajzolódása vonatkozásában. Az online közösségi háló térkapcsolataiból ugyanis várható volt, hogy kiolvashatók azok a települési hierarchiai központok, amelyek a társadalmi térszerveződésben valódi hierarchiai centrumokként értelmezhetők. Ezek az eredmények összevethetők a közigazgatási térszerveződés centrumaival is, azaz megvizsgálható, hogy a közigazgatási értelemben vett centrumok és a spontán társadalmi kapcsolatokban realizálódó centrumok egybe esnek-e. Persze nem kell hozzá nagy fantázia, hogy lássuk, hogy a felhasználói létszámok és a kapcsolatsűrűség a települési népességszámmal valószínűleg erősen korrelálnak, de mégis feltételezhetők ettől való eltérések, illetve az eddigi településközi kapcsolatsűrűségekre vonatkozó sejtéseink végre igazolódhatnak.

A kapcsolatok száma alapján felvázolható ábra nem meglepő módon a legnépesebb településeink között mutatja a legsűrűbb összeköttetéseket (3. ábra). Ahogy vártuk, a településközi abszolút kapcsolatszám egyértelműen a településhierarchia mentén rendeződik Magyarországon, így a térképen hub-ként jelennek meg a regionális központok, s természetesen a legjelentősebb hub-nak Budapest adódott. A modell eredményei visszaigazolták az igazgatási hierarchia hazai főbb gócpontjait.



3. ábra

A településhierarchia tesztelése: a legsűrűbb településközi kapcsolatok az iWiW közösségi hálóban (2013; csak a 15 000 db feletti kapcsolatszámú összeköttetéseket jelölve)

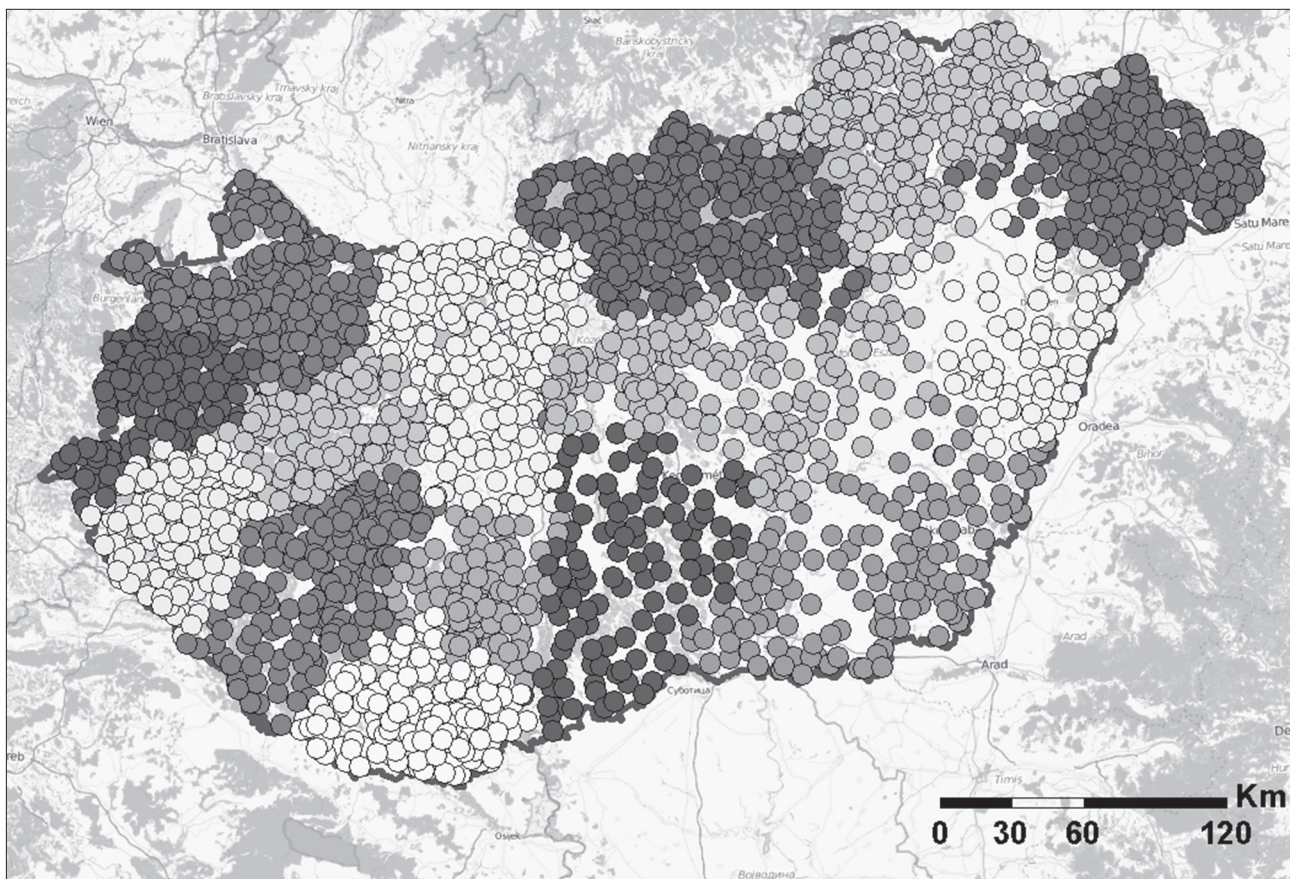


4. ábra

A szoros térkapcsolatok tesztelése: a legerősebb súlyú településközi kapcsolatok az iWiW közösségi hálóban (2013)

Az abszolút kapcsolatszámokat megjelenítő ábra ugyanakkor nem képes tájékoztatást adni a kapcsolatok szorosságáról, azaz nem tükrözi, hogy az egyes települések kapcsolati szerkezetében melyek a fontos partnerek, az erős szálakkal összekötött esetek, és melyek a gyenge kapcsolatok, pedig az igazgatási térfelosztás tesztelése szempontjából épp ezen erős összeköttetések lennének a meghatározók. A hálózati modellben tehát azokat a településközi kapcsolatokat kellene megtalálni, amelyek nemcsak egy-egy felhasználó, de mindenki számára fontosak. Annak érdekében tehát, hogy a virtuális térben létező szoros kapcsolatokat is feltárhassuk, a települések közti összeköttetéseket minden település esetében a fontosságuk szerint súlyozni kellett. A kialakított log-normalizált kapcsolatsúly eredmények így lehetővé teszik a fontos kapcsolatok kihangsúlyozását a településmérettől függetlenül is. A kapcsolatsúly a megfigyelt és a véletlenszerűen várható településközi kapcsolatszám egymáshoz viszonyított arányaként került meghatározásra (részletesebb leírását lásd Jakobi – Lengyel 2015).

A súlyozatlan, nyers adatokhoz képest a súlyozott kapcsolati adatok alapján kirajzolódó térkép már egészen más rajzolatot mutat (4. ábra). A legerősebb kapcsolati súlyérték-



5. ábra

A horizontális térfelosztás tesztelése: modularitás az iWiW közösségi hálóban (2013)

kekkel bíró élek nem a legnagyobb települések között húzódnak, hanem inkább a kisebbek, és főként a közel lévők között (bár egyes helyeken azért természetesen láthatók kivételek). A relatív élsúlyokat megjelenítő térkép csak a 4,0 feletti kapcsolati súllyal rendelkező összeköttetéseket mutatja, mivel e limit felett valóban csak a legerősebb kapcsolatok figyelhetők meg, s azért is, mert az ábrázolandó vonalak mennyisége így még kezelhető és áttekinthető. Az ábra összességében 34968 összeköttetést jelenít meg, ezzel érzékelte a legfontosabb helyi kapcsolatok területi szerkezetét. A kis távolságú, de erős és szignifikáns virtuális térbeli kapcsolatok így akár az igazgatási mikroterek kapcsolati rendszereit is képesek viszsztatükrözni.

A hálózati modellek egyik potenciálisan feltárható sajátossága a hálózaton belül kialakuló alhálózatok vagy erősebb kapcsolódású csoportok meghatározása. Az online közösségi hálóban földrajzilag azonosítható klaszteresedés szintén jelen lehet, melynek eredménye a hálózat térbeli modularitása, ami ha így van, a hálózat viszonylag összefüggő területi egységekre tagolódását jelzi (Thiemann et al 2010). Annak érdekében, hogy a klasztereket azonosíthassuk a városhálózatban, a közösségek azonosítására alkalmas Louvain módszert alkalmaztuk (melynek részletesebb leírását lásd Lengyel et al. 2015, 2016). Az eredményül kapott térben meghatározott csoportok a térszerveződés összefüggő egységeit jelölik, amelyek bár a virtuális térbeli kapcsolataink adataiból épül-

nek fel, de mégis a valós világunk igazgatási térszerkezetéből ismert horizontális térfelosztás mintázatát követik többé-kevésbé, ami meglepő eredmény.

Az 5. ábrán látható, hogy a gráfok közösségi struktúrája területi alapon, tehát nem térben véletlenszerűen jött létre. A súlyozott kapcsolati adatokon lefutott hálózati modularitás vizsgálatok a Louvain algoritmus 14 klasztert hozott létre. A hálózat klaszterei néhány kivételtől eltekintve a megyehatároknak (NUTS-3 szint) felelnek meg. Ezek a közigazgatási egységek (a történelmi hagyományoktól nem függetlenül) 1950-ben lettek létrehozva és évtizedeken keresztül a gazdasági és társadalmi tervezés alapegységei voltak. A Budapest körül lévő agglomerációs területek nem képeznek önálló egységet, hanem a hozzájuk közel eső klaszterek részeivé válnak. Érdekes megjegyezni azt is, hogy korábbi eredményekhez hasonlóan a Duna választóvonalát itt is megfigyelhető (Mate et al 2011), azaz a folyó a Louvain algoritmus futtatása során határként szolgált a közösségek számára is. Persze a moduláris térkép nem tökéletesen illeszkedik a közigazgatás jelenlegi megyei szintű területi felosztásához, azaz egyes modulok több megyét is lefednek (lásd pl. Csongrád-Békés, Nógrád-Heves vagy Fejér-Komárom-Esztergom példáit), máshol pedig új, de talán indokolható „megyék” is kialakultak (lásd például a Balaton környékén kirajzolódó önálló klasztert). Ez a módszer így újabb adalékokkal szolgálhat egy esetleges új területi közigazgatási beosztás kialakításához.

4. Összegzés

Vitathatatlan, hogy a big data jelenség napjaink egyik legfelkapottabb témája, s nem véletlenül. Se szeri, se száma azoknak az alkalmazási területeknek, ahol a big data állományokban rejlő lehetőségeket mostanában kezdik felismerni, sőt kihasználni is. Ezen a fronton a területi tervezés és a területi igazgatás döntés-előkészítő tevékenységei sem maradhatnak le. Az újszerű területi adatok felhasználási lehetőségei a várostervezés, a közlekedéstervezés, a turisztikai fejlesztések, a területfejlesztés vagy a területi igazgatásszervezés szakterületein nemcsak új látásmóddal, de új eredményekkel is kecsegtetnek.

Irodalom

- BENEDIKT, M. (1991): *Cyberspace: some proposals*. In Benedikt, M. (ed.) *Cyberspace: first steps*. MIT, Cambridge, 119–224.
- BOYD, D., ELLISON, N.B. (2007): *Social Network Sites: definition, history, and scholarship*. Journal of Computer Mediated Communication 13, Article 1. (Letöltés: 2012.11.20.) <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>
- CARDEN, T., COAST, S. (2005): *London GPS tracking map*. OpenStreetMap. <http://www.visualcomplexity.com/vc/project.cfm?id=256>
- ELLISON, N., STEINFELD, C., LAMPE, C. (2006): *Spatially Bounded Online Social Networks and Social Capital: The Role of Facebook*. Paper to be presented at the Annual Conference of the International Communication Association, June 19-23, 2006 in Dresden, Germany.
- FERNBACK, J. (2007): *Beyond the diluted community concept: a symbolic interactionist perspective on online social relations*. New Media Society, 9., 1., 49–69.
- FISCHER, E. (2013): *Locals and Tourists map*. Gnip, MapBox project. Elérhető: <http://mapbox.com/labs/twitter-gnip/locals>
- FRIEDEWALD, M., RAABE, O. (2011): *Ubiquitous computing: An overview of technology impacts*. Telematics and Informatics 28. pp. 55–65.
- GALLOWAY, A. (2004) *Intimations of everyday life: Ubiquitous computing and the city*. – Cultural Studies 18. (2) pp. 384–408.
- GICZI J., SZÓKE K. (2017): *Hivatalos statisztika és a Big Data*. Statisztikai Szemle, 5. (95), pp. 461–490.
- GIRARDIN, F., VACCARI, A., GERBER, A., BIDERMAN, A., RATTI, C. (2009): *Quantifying urban attractiveness from the distribution and density of digital footprints*. International Journal of Spatial Data Infrastructures Research 4., pp. 175–200.
- HECHT, B., HONG, L., SUH, B., CHI, E. H. (2011): *Tweets from Justin Bieber's heart: the dynamics of the „location” field in user profiles*. Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2011), ACM Press, New York, 237–246.
- HELLERSTEIN, J. (2008): *The Commoditization of Massive Data Analysis*. University of California, Berkeley, USA. <http://radar.oreilly.com/2008/11/the-commoditization-of-massive.html>
- JAKOBI Á., LENGYEL B. (2015): *Geovisualising unequal spatial distribution of online social network connections: a Hungarian example*. In: Brus J, Vondrakova A, Vozenilek V (szerk.) *Modern Trends in Cartography. Lecture Notes in Geoinformation and Cartography series*, Springer International Publishing, Heidelberg – New York, pp. 227–240.
- JÄRV, O., AHAS, R., SALUVEER, E., DERUDDER, B., WITLOX, F. (2012) *Mobile Phones in a Traffic Flow: A Geographical Perspective to Evening Rush Hour Traffic Analysis Using Call Detail Records*. PLoS ONE 7 (11) e49171. doi:10.1371/journal.pone.0049171
- JIANG, B., YAO, X. (2006): *Location-based services and GIS in perspective*. – Computers, Environment and Urban Systems 30. pp. 712–725.
- JONES, S. G. (1995): *Understanding community in the information age*. In Jones, S. G. (ed.) *Cybersociety: computer-mediated communication and community*. Sage, London, 10–35.
- KING, G. (2011): *Ensuring the Data-Rich Future of Social Science*. Science, 331, pp. 719–721.
- LENGYEL B., VARGA A., SÁGVÁRI B., JAKOBI Á., KERTÉSZ J. (2015): *Geographies of an online social network*. PLoS ONE 10(9): e0137248. doi:10.1371/journal.pone.0137248
- LENGYEL B., VARGA A., SÁGVÁRI B., JAKOBI Á., KERTÉSZ J. (2016): *Az iWiW földrajza*. Területi Statisztika, 56(1), (megjelenés alatt), p. 17.
- MATE G, NEDA Z, BENEDEK J. (2011): *Spring-Block model reveals region-like structures*. PLoS ONE; 6: e16518.
- NAAMAN, M. (2011): *Geographic information from georeferenced social media data*. SIGSPATIAL 3 (2) pp. 54–61.
- NHIT (2016): *Fehér Könyv a nemzeti adatpolitikáról*. Nemzeti Hírközlési és Informatikai Tanács Szakértői Tanácsadó Testülete, 2016. július, Budapest.
- ONNELA, J. P. (2011): *Social Networks and Collective Human Behavior*. UN Global Pulse. <http://www.unglobalpulse.org/node/14539>
- PINTÉR R. (2011): *Álmodnak-e a telefonok elektronikus bárányokkal?* Konferencia-előadás. iPhone vs. Android-konferencia. 2010. április 8., Budapest. <http://www.slideshare.net/hirbehozo/pinter-rbert-mobil-trendek>
- SATYANARAYANAN, M. (2001): *Pervasive computing: vision and challenges*. Personal Communications, IEEE 8. (4.) pp. 10–17.
- SZÜTS Z., YOO, J. (2016): *Big Data, az információs társadalom új paradigmája*, Információs Társadalom, XVI. évf. (2016) 1. szám, 8–28. old.
- TAKHTEYEV, Y., GRUZD, A., WELLMAN, B. (2012): *Geography of Twitter networks*. Social Networks, 34, 73–81.
- THIEMANN C, THEIS F, GRADY D, BRUNE R, BROCKMANN D, (2010) *The structure of borders in a small world*. PLoS ONE, 5: e15422.
- WEISER, M. (1991): *The computer for the 21st century*. Scientific American, 265(3) pp. 94–104.
- ZOOK, M., DODGE, M., AOYAMA, Y., TOWNSEND A. (2004): *New Digital Geographies: Information, Communication, and Place*. In. Brunn, S.D. - Cutter, S.L. – Harrington, J. W. (eds.) *Geography and Technology*. Kluwer Academic Publishers. pp. 155–176.