



Közzététel: 2023. december 12.

A tanulmány címe:

Az elvándorlás mérésének digitális lehetőségei a statisztikában

Szerzők:

T. NAGY JUDIT

a Károli Gáspár Református Egyetem docense; E-mail: tnagy.judit@kre.hu

BALOGH ESZTER

a Károli Gáspár Református Egyetem docense; E-mail: balogh.eszter@kre.hu

CZIRÁKI KÁROLY TAMÁS

a Károli Gáspár Református Egyetem hallgatója; E-mail: czirakikarcsi@gmail.com

TAKÁCS KAROLINA FRUZSINA

a Károli Gáspár Református Egyetem hallgatója; E-mail: takacs.kfr@gmail.com

CSANÁDY MÁRTON

a Károli Gáspár Református Egyetem tudományos főmunkatársa; E-mail: csanady.marton.tamas@kre.hu

SZVETELSZKY ZSUZSANNA

a Károli Gáspár Református Egyetem docense; E-mail: szvetelszky.zsuzsanna@kre.hu

DOI: <https://doi.org/10.20311/stat2023.11.hu1084>

Az alábbi feltételek érvényesek minden, a Központi Statisztikai Hivatal (a továbbiakban: KSH) Statisztikai Szemle c. folyóiratában (a továbbiakban: Folyóirat) megjelenő tanulmányra. Felhasználó a tanulmány vagy annak részei felhasználásával egyidejűleg tudomásul veszi a jelen dokumentumban foglalt felhasználási feltételeket, és azokat magára nézve kötelezőnek fogadja el. Tudomásul veszi, hogy a jelen feltételek megszegéséből eredő valamennyi kárért felelősséggel tartozik.

1. A jogszabályi tartalom kivételével a tanulmányok a szerzői jogról szóló 1999. évi LXXVI. törvény (Szjt.) szerint szerzői műnek minősülnek. A szerzői jog jogosultja a KSH.
2. A KSH földrajzi és időbeli korlátozás nélküli, nem kizárólagos, nem átadható, térítésmentes felhasználási jogot biztosít a Felhasználó részére a tanulmány vonatkozásában.
3. A felhasználási jog keretében a Felhasználó jogosult a tanulmány:
 - a) oktatási és kutatási célú felhasználására (nyilvánosságra hozatalára és továbbítására a 4. pontban foglalt kivétellel) a Folyóirat és a szerző(k) feltüntetésével;
 - b) tartalmáról összefoglaló készítésére az írott és az elektronikus médiában a Folyóirat és a szerző(k) feltüntetésével;
 - c) részletének idézésére – az átvevő mű jellege és célja által indokolt terjedelemben és az eredetihez híven – a forrás, valamint az ott megjelölt szerző(k) megnevezésével.
4. A Felhasználó nem jogosult a tanulmány továbbértékesítésére, haszonszerzési célú felhasználására. Ez a korlátozás nem érinti a tanulmány felhasználásával előállított, de az Szjt. szerint önálló szerzői műnek minősülő mű ilyen célú felhasználását.
5. A tanulmány átdolgozása, újra publikálása tilos.
6. A 3. a)–c) pontban foglaltak alapján a Folyóiratot és a szerző(ke)t az alábbiak szerint kell feltüntetni:
„*Forrás: Statisztikai Szemle c. folyóirat 101. évfolyam 12. számában megjelent, T. Nagy Judit–Balogh Eszter–Cziráki Károly Tamás–Takács Karolina Fruzsina–Csanády Márton–Szvetszky Zsuzsanna által írt, Az elvándorlás mérésének digitális lehetőségei a statisztikában című tanulmány (link csatolása)*”
7. A Folyóiratban megjelenő tanulmányok kutatói véleményeket tükröznek, amelyek nem feltétlenül esnek egybe a KSH vagy a szerzők által képviselt intézmények hivatalos álláspontjával.

**T. Nagy Judit – Balogh Eszter – Cziráki Károly Tamás – Takács Karolina Fruzsina
– Csanády Márton – Szvetselszky Zsuzsanna**

Az elvándorlás mérésének digitális lehetőségei a statisztikában

Digital methods for measuring migration in statistics

T. Nagy Judit, a Károli Gáspár Református Egyetem docense
E-mail: tnagy.judit@kre.hu
Balogh Eszter, a Károli Gáspár Református Egyetem docense
E-mail: balogh.eszter@kre.hu
Cziráki Károly Tamás, a Károli Gáspár Református Egyetem hallgatója
E-mail: czirakikarcsi@gmail.com
Takács Karolina Fruzsina, a Károli Gáspár Református Egyetem hallgatója
E-mail: takacs.kfr@gmail.com
Csanády Márton, a Károli Gáspár Református Egyetem tudományos főmunkatársa
E-mail: csanady.marton.tamas@kre.hu
Szvetselszky Zsuzsanna, a Károli Gáspár Református Egyetem docense
E-mail: szvetselszky.zsuzsanna@kre.hu

Azokra a hagyományos „offline” adatforrásokra, amelyekből meghatározható a vizsgált társadalmi csoport mérete, az adatgyűjtés természetéből adódóan nem jellemző, hogy egyszerre megbízhatóak és időszerűek is. Ezek az adatforrások azonban jól kombinálhatók „online” társaikkal, amelyeken keresztül a digitális térben történő társadalmi interakciók nyomán keletkező és eredetileg nem kutatási célból gyűjtött adatok érhetőek el. Az online adatoknak azon része, amely a közösségimédia-portálok használata során keletkezik, könnyen hozzáférhető, időszerű, magas szintű lebontással rendelkezik, és a közösségimédia-portálok kiterjedtségének köszönhetően széles lefedettséggel bír. Tanulmányunkban egy példán keresztül mutatjuk be, hogy a kétféle adatforrás kombinálásával hogyan készíthető valós idejű becslés az Ausztriában élő magyar nők számáról.

Kulcsszavak: digitális demográfia, női migráció, mobilitás, közösségi média, Facebook

The traditional 'offline' data sources, which can be used to measure the size of the migrant social group can not be timely and reliable at the same time. These data sources can be combined well with their 'online' counterparts, which are not originally collected for research purposes but as they result from social interactions in the digital space can be accessed also for the researchers. The online data that is generated through the use of social media portals is easily accessible, timely, has a high level of disaggregation and has a wide coverage due to the extent of social media portals. In our study, we use an example to show how combining these two data sources can produce a real-time estimate of the number of Hungarian women living in Austria.

Keywords: digital demography, female migration, mobility, social media, Facebook

Jelen tanulmány bemutatja, hogy a digitális demográfiában is alkalmazott Big Data-módszertannal hogyan lehet az elvándorló társadalmi csoport méretéről időszerű és demográfiailag részletes adatokhoz jutni. A módszertan a közösségi média, azon belül a Facebook adatait használja fel és kombinálja a rendelkezésre álló offline adatokkal. A cikk a módszert az Ausztriába elvándorló magyar nők példáján keresztül ismerteti.

1. A migrációs adatokkal kapcsolatos dilemmák

Sajnálatos módon a migrációról pontos adatok sem a politikai döntéshozók, sem a tudományos világ számára nem állnak rendelkezésre. Számptalan hazai (Sik, 2013; Golovics, 2019; Hárs–Sik, 2009) és nemzetközi tanulmány (Bijak–Czaika, 2020; Willekens et al., 2016) született arról, hogy a kibocsátó és a fogadó országok offline adatai miért nem alkalmasak a populáció pontos méretének a becslésére, illetve a regisztrációs adatok miért nem szolgálnak megbízható adatforrásként. A kibocsátó országok statisztikái általában sokkal kevesebb migráns (külföldön élő) egyénről számolnak be, mint a fogadó országok tükörstatisztikái (Golovics, 2019), ráadásul a nemzetállamok statisztikai adatait sem lehet mindig összehasonlítani egymással. (A konceptualizációs problémákból eredő kérdéseket Golovics {2019} tárgyalja részletesen.) A jelenleg használt adatforrások tehát a legtöbb esetben következtelenek és pontatlanok, illetve az adatgyűjtés és az adatok terjesztése között jelentős idő telik el (Abel–Sander, 2014; Willekens et al., 2016).

Fontos kiemelni a módszerek „elavultságát” is. A migrációkutatás hagyományos, múlt századi fogalmi és módszertani rendje, amely a jelenség egyirányúságából és a letelepedés tartósságából indul ki, ma már kevésbé alkalmas a vándorlási mozgalmak leírására (Ligeti, 2021). A vándorlás ma már oda-vissza történő mozgások összessége, és nem feltétlenül jár együtt tartós letelepedéssel. Többek között Ligeti (2019, 2021) kutatásai hívták fel a figyelmet a magyar migráció likvid (Engbersen–Snel, 2013) jellegére, ami keresztmetszeti helyett longitudinális elemzést igényel (Ligeti, 2021). A migrációs trendek előrejelzésére és a migrációban élő populáció becslésére több módszerrel is próbálkoztak már, közülük fontos kiemelni Sik Endre „migrációs potenciál” fogalmát (2003; 2018), amellyel

javasolja a mozgás előrejelzéséhez a longitudinális kérdőíves adatfelvételeket. A próbálkozások ellenére mégis az a tapasztalat, hogy a migrációban élő populáció méretének pontos és valós idejű becslése, demográfiai jellemzőinek megismerése sok módszertani és fogalmi buktatóval jár.

A fentiek alapján úgy gondoljuk, hogy a migrációs trendek predikciójára a hagyományos adatfelvételi módszerek kevésbé alkalmasak, mivel csak korlátozott hatékonysággal tudják a mozgás dinamikáját és „tömegét” megfigyelni. Mindeközben egyre fontosabbá vált, hogy a migránspopulációról megfelelő adatok álljanak rendelkezésre. Jelen tanulmány bemutatja, hogy a „hagyományos keretek közül” átlépve az online, digitális térben kialakuló adatok világába, milyen új módszerek jelentek meg a tudományban a nemzetközi mozgás predikciójára és a migrációban élő populáció becslésére.

2. Digitális demográfia és digitális migrációkutatás

Cesare és munkatársai (2015), valamint *Bohon (2018)* egyértelműen állást foglalnak amellyel, hogy a demográfiai kutatásoknak az analóg, hagyományos keretből ki kell lépniük és a Big Data-adatok felé kell fordulniuk. Javasolják az új adatfelvételi módszerek és az ehhez szükséges adatvédelmi gyakorlatok, elvek kidolgozását. Úgy gondolják, hogy a demográfia módszertanának a digitális térbe történő belépése elkerülhetetlen a módszertan evolúciója során. Elképzelésüket arra alapozzák, hogy a digitalizáció térnyerésével az emberek jelentős mennyiségű digitális lábnyomot (*Latour, 2007*) hagynak maguk után. A digitális nyomok a digitális eszközökön és tereken keresztül megvalósuló társadalmi interakciók eredményei, valamint különféle anyagok digitális nyilvántartásai és az emberi viselkedés digitális dokumentálásának formái (például a digitális eszközökön keresztül zajló interakciók). Digitális nyomok például az e-mailek, a hitelkártya-tranzakciók, a mobiltelefon-nyilvántartások, a közösségimédia-adatok.

2.1. Digitális adatgyűjtés a migrációkutatásban

A digitális nyomok eredetileg nem kutatási célból gyűjtött adatok (*Cesare et al., 2018*), de a kutatók számára kiváló források. Ahogyan a népességet érintő folyamatok, úgy a migráció is vizsgálható digitális nyomkövetési adatok felhasználásával. A migrációs adatokkal kapcsolatos problémák kezelésének az igénye és

a digitális demográfia berobbanása vezetett oda, hogy a migrációs kutatásokban is új adatfelvételi trendek jelentek meg.

A közösségi média, az e-mailek és a mobiltelefonok kiváló adatforrásként szolgálnak a vándorlási mozgalmak megfigyelésére. *Zagheni és Weber (2012)* például a kor- és nemspecifikus migrációs arányokat az e-mailek nagymintájából nyert adatok alapján becsülték meg. Megállapították, hogy a migránsok életkori profilú becslése közel állt a meglévő adminisztratív adatokhoz. Azt a következtetést vonták le, hogy az e-mail-adatok érdemben egészíthetik ki a rendelkezésre álló migrációs adatokat és közelebb vihetnek a mozgás mintáinak és a hálózatoknak a megértéséhez.

Massicotte és Eddelbuettel (2020) a migrációval kapcsolatos keresési kulcsszavakat gyűjtötte össze. Eredményeik azt mutatták, hogy a munkakereséshez és az oktatáshoz kapcsolódó keresési kulcsszavak összetett indexei javíthatják az egy évre előre mutató előrejelzéseket, amelyek kiegészíthetik a hivatalos migrációs statisztikákat, és így segíthetnek a migrációs áramlások tendenciáiban bekövetkező esetleges hirtelen változások felderítésében. Az összes kulcsszóból összeállított Google Trends-index viszonylag kis mértékű eltérést mutatott a megfigyelt migrációs áramlásokhoz képest. Fontos korlátot jelent azonban, hogy a keresőmotorok adatainak felhasználása az előrejelzett migrációs áramlások jellemzőitől függ. *Abel–Sander (2014)* a Google Correlate (*Mohebbi et al., 2011*) és a Google Trends adatait használta a munkaerő mobilitásának mérésére Ausztráliában. *Gendronneau és munkatársai (2019)* a Google Trends által hivatalosan közzétett adatokat egészítették ki a Facebook hirdetési platformjáról származó információkkal. Kutatásuk célja az volt, hogy rövid távú és valós idejű előrejelzéseket közöljenek a migráns populációról.

2.2. A közösségi portálok mint adatforrások

A közösségimédia-platformok azért jöttek létre, hogy a felhasználók nyilvános vagy részben nyilvános profilokon keresztül kapcsolatba léphessenek egymással, online közösségeket létesíthessenek, tartalmat hozhassanak létre és oszthassanak meg. A felhasználóknak létrejön egy saját adatlapjuk/profiljuk/csatornájuk, amelyen a saját maguk által megosztott tartalmak (például üzenet, kép, videó vagy esemény) jelennek meg. Követéssel, ismerőssé válással a felhasználók kapcsolódhatnak egymáshoz, így láthatják egymás tartalmát, hírfolyamát (*feed*), és reagálhatnak is arra. Emellett a hírfolyamban a felhasználók számára személyre szabott tartalmak is megjelennek.¹

¹ A személyre szabott tartalmakat az alkalmazás a profilozó algoritmusával segítségével biztosítja a felhasználó által megnézett, reagált tartalmak és a geolokáció alapján.

2.2.1. Látható tartalmak

A közösségimédia-platformokon a felhasználók tevékenységei során keletkező digitális nyomok egy része hozzáférhető és kutatási célokra felhasználható a cikk írásának időpontjában.² Az oldalak látható tartalma elérhető, és akár manuális úton, akár web scraper botokkal letölthető.

Kádár és Gede (2021) kutatásuk során a Flickr és az Instagram adatait (geotaggel rendelkező fényképeket, a feltöltés időpontjával) használták fel a láthatatlan turizmus vizsgálatára, és statisztikailag is releváns szabadidő-használati mintázatokat tártak fel. *Zagheni és munkatársai (2014)* a Twitter adatai alapján kutatták a rövid távú migrációt, a Twitter-felhasználók dátummal és geotagadatokkal rendelkező tweetjein keresztül tanulmányozták a nemzetközi és a belső migrációt az OECD-országokban.

2.2.2. Személyes adatok

A legjelentősebb közösségimédia-platformok többsége lehetőséget biztosít arra, hogy az egyes felhasználók a saját fiókadataik másolatát bármikor megtekinthessék, vagy akár le is töltsék. Ezek az adatok tartalmazzák a Facebook-tevékenységeket (például a hozzászólásokat, a kedveléseket, a kereséseket, a bejelentkezéssel kapcsolatos információkat, a hirdetési információkat, az ismerőssé válásokat), valamint az adatlapon megadott (akár nyilvános, akár rejtett) adatokat. E lehetőséget használta ki például *Kmetty és Bozsonyi (2022)*, valamint *Kmetty és Németh (2022)* is, akik Facebook-felhasználókat kértek meg az adatprofil-archívumuk exportálására és megosztására. Az első szerzőpáros az így kapott adatokat kérdőíves adatgyűjtéssel kombinálva megkísérelte a depressziós viselkedés digitális lenyomatait beazonosítani. A másik szerzőpáros pedig a saját bevallású adatok, valamint a digitálisan kifejezett érdeklődés és a Facebook profilozó algoritmusai által meghatározott érdeklődés összevetésére irányult, azok kapcsolatát vizsgálta.

2.2.3. Aggregált adatok

A közösségi portálok üzleti modellje a hirdetéseken alapul. Bizonyos portálok lehetővé teszik a hirdető számára, hogy a felhasználók társadalmi-demográfiai jellemzőit szűrőfeltételként használva – a földrajzi részletesség különböző szintjein összesítve –, a feltételeknek eleget tevő felhasználók (célközönség) becsült számát lekérdezzék, majd díjfizetés után megcélozhatják őket hirdetéseikkel.

² A hozzáférhetőség folyamatosan változik, hiszen a kezdeti időszakhoz képest a szolgáltatók mára már zárt rendszereket alakítottak ki és csak saját felületükön keresztül hozzáférést engedélyeznek. A portálok megalakulásakor még nyilvános hozzáférésű API-val elérhetőek és szisztematikusan kereshetőek voltak az adatbázisok.

Ezeket a funkciókat a kutatók kétféleképpen is kihasználhatják. Egyrészt maguk a célközönségbecslések is hasznos információkat szolgáltatnak, másrészt a közösségimédia-populáció mintavételi keretként is szolgálhat, hiszen egyes csoportok megcélozhatók olyan hirdetésekkel, amelyeknek célja, hogy bevonják őket egy adott felmérésbe. Ily módon elérhetővé válnak a hagyományos eszközökkel nehezen elérhető társadalmi csoportok. *Pöttschke és Braun (2016)* például a nehezen elérhető lengyel migránsokról gyűjtött adatokat, olyan egyénekről, akiket hagyományos kutatási módszerekkel nehezen lehet elérni. A kutatók Facebook-hirdetések révén vettek mintát lengyel migránsok adataiból Ausztriában, Írországban, Svájcban és az Egyesült Királyságban, majd utólagos rétegzési technikák segítségével súlyozták a válaszadókat, és vontak le az alapsokaságra vonatkozó statisztikai következtetéseket.

2.2.4. Előnyök és az adatok használhatóságának korlátai

A közösségi portálokról kinyerhető digitális nyomok tehát új lehetőséget jelentenek a migráció vizsgálatában, és a következőképpen jellemezhetők.

1. A közösségi médiahasználat elterjedtségének köszönhetően széles lefedettséggel bírnak.
 - a. Megfelelő statisztikai módszerekkel keresztmetszeti adatokhoz juthatunk a migrációs populáció méretére és demográfiai összetételére vonatkozóan.
 - b. Lehetővé teszik a kutatók számára, hogy hozzáférjenek olyan csoportokhoz, amelyeket nehéz elérni,³ és/vagy a hagyományos felmérési technikákkal általában alulreprezentáltak (*Zagheni–Weber, 2012; Zagheni et al., 2014*) (*Cesare et al., 2018; Zagheni et al., 2017*).
2. Magas térbeli lebontás jellemző rájuk, így például a geotagek használatával a hagyományos adatforrásokhoz képest pontosabb adatokat szolgáltathatnak.
3. A hagyományos adatforrásokhoz képest plusz információt (például demográfiai és viselkedési adatokat és érdeklődési köröket) is szolgáltathatnak.
4. Valós idejűek, gyorsan elérhetőek, emiatt longitudinális vizsgálatokra, a migrációs trendeknek és a migráció dinamikájának a vizsgálatára tökéletesen alkalmasak, így segítségükkel a népesség vándorlásának mögöttes magyarázatát is megérthetjük.

Az előnyökön túl azonban nem szabad megfeledkeznünk bizonyos korlátokról sem. A közösségimédia-portálokról gyűjtött adatok teljes lakosságra történő általánosíthatóságával kapcsolatban az a legfőbb probléma, hogy a portálok adatot

³ Például nem áll rendelkezésre nyilvántartás vagy mintavételi keret.

szolgáltató felhasználói nem képviselik reprezentatívan a teljes lakosságot. Ennek az az oka, hogy a portálok elterjedtsége a népesség különböző kategóriái (ország, életkor, nem) között eltérő lehet (például az idősebbek jelenléte korlátozottabb, mint a fiataloké). Ez a probléma torzításhoz vezethet a teljes népességre vonatkozó becslések során. Ennek a reprezentativitási torzításnak a korrigálására többféle módszer is alkalmazható (*Spyratos et al., 2018; Zagheni et al., 2014; Zagheni–Weber, 2015*). A módszerek alapelve, hogy a közösségimédia-portálok segítségével becsült számadatok a megfelelő hivatalos adatok, valamint a penetrációs-ráta-adatok felhasználásával kalibrálhatók.

A reprezentativitási torzítás részét képezi az is, hogy a portálok csak azokról szolgáltatnak bizonyos információkat, akik rendszeres internetkapcsolattal rendelkeznek (*Bohon, 2018*). További probléma, hogy bizonyos adatok önbevalláson alapulnak, így torzítást okozhatnak az önmagukat hamisan ábrázoló egyének (például a regisztráció érdekében önmagukat szándékosan idősebbnek beállító felhasználók, vagy az egy-egy személyhez tartozó, nem feltétlenül szándékosan duplikált profilok) (*Cesare et al., 2018*), a hiányos adatlapok (*Bohon, 2018*)⁴, valamint a botoktól származó kéretlen adatok is (*Cesare et al., 2018*).

Ugyanakkor, amennyiben a számítási modellben ezeket, a becslést torzító tényezőket és más, akár az adott kibocsátó országra jellemző sajátosságokat is figyelembe vesszük, a torzítás mértéke jelentősen csökkenhető. Végül az is korlátot jelenthet, hogy a közösségimédia-portálokról kinyerhető adatok csak valós idejűek lehetnek, múltbeli adatok nem érhetők el.

2.2.5. A Facebook Hirdetéskezelő alkalmazása a migrációs kutatásokban

A migrációval kapcsolatos adatfelvételek kiemelkedő színtere a Facebook, mivel demográfiaileg sokszínű populáció használja (*Barberá, 2016; Lewis et al., 2013; idézi: Cesare et al., 2018*) és közel 2 milliárd felhasználójával jelenleg a legnagyobb közösségimédia-platformnak tekinthető.

A Facebook Hirdetéskezelő (*Adverts Manager Interface/Marketing API*) segítségével – ahogy az előző alfejezetben leírtuk – megbecsülhető azoknak a felhasználóknak (profiloknak) a száma, akik bizonyos szűrőfeltételeknek eleget tesznek. A Facebook Hirdetéskezelő adatai a cikk írásának időpontjában bárki számára ingyenesen lekérhetőek üzleti Facebook-fiók birtokában. Az adatok elérése Facebook Hirdetéskezelő Interface-en vagy a Facebook Marketing API-n keresztül valósítható meg.

A kiválasztható szűrőfeltételek közé tartoznak a Facebook-felhasználók által kifejezetten megadott információk (például nem, életkor, lakóhely, születési

⁴ A hiányos adatlap a magyar felhasználók fontos sajátossága.

hely), valamint a Facebookon és a kapcsolódó webhelyeken történt felhasználói interakciókból kikövetkeztetett információk (például érdeklődési körök, viselkedések).⁵ Az óriási mennyiségű összesített anonim adaton felül tehát az adatok minősége is különleges, hiszen Facebook Hirdetéskezelő a látható (és nem látható) profiladatokon túl az online magatartásról is gazdag adatmennyiséget biztosít.

A migrációval foglalkozó kutatók számára kiemelkedő jelentőségű, hogy a Facebook – az adatlapon megadott születési hely, korábbi és jelenlegi lakóhelyek, valamint a bejelentkezési helyek nyomon követése (véltetően IP-cím, geotagek) alapján – rendelkezésre bocsátja szűrőként az alábbi tulajdonságokat (<https://business.facebook.com/adsmanager/>):

- jelenlegi lakóhely („Olyan emberek, akiknek az otthona a megadott területen belül van”);
- külföldön élők (korábban: expatok) („Olyan emberek, akik korábban [...] országban éltek, de most más országban élnek”), oly módon, hogy a Facebook nyilvántartja a korábbi lakóhelyeket;
- ingázók („Olyan emberek, akik valószínűleg hétköznapiokon ingáznak a munkahelyükre”).

Ezen felül szűrhetünk a nyelvre is, azaz olyan felhasználókra, akik egy adott „földrajzi helyen nem jellemző nyelvet használnak”. A fenti szűrők a Facebook által kialakított mesterséges változók, a hozzájuk megadott definíciók az idézőjelekben láthatók.

A Facebook dokumentációja azonban nem ad felvilágosítást arról, hogy milyen egyéni jellemzőkből és hogyan történik ezen változók létrehozása. A külföldön élők (expatok) kategóriát például több tanulmányban (*Herdağdelen et al., 2016; Herdağdelen–Marelli, 2017; Spyrtos et al., 2019*) is vizsgálták, különböző eredménnyel. Ezenkívül 2018-ban módosítás történt a definícióban, akkor valószínűleg megváltozott a felhasználót migránsként azonosító algoritmus is. A Facebook „Facebook Inc. 2019 Annual Report 10-K” című hivatalos dokumentumában további információ található erről: „a felhasználók földrajzi elhelyezkedése számos tényező alapján kerül megbecsülésre, mint például a felhasználók IP-címe és saját maguk által közölt tartózkodási helye” (*U.S. SEC, 2020*).

További problémát jelenthetnek a fogalmi eltérések, például a Facebook „külföldön élők” és a hivatalos statisztikák „migráns” meghatározása között (a Facebook által nyilvántartott „szülőváros” országa nem biztos, hogy azonos az Eurostat nyilvántartásában szereplő születési országgal). Ezen túlmenően a Facebook fix változókört és az aggregáláshoz fix kategóriákat használ, amelyek nem biztos, hogy egyeznek a kutatási igényünkkel vagy a kalibráláshoz felhasznált

⁵ A szűrőfeltételek változóit a Facebook rendszeresen frissíti.

hivatalos adatforrások kategóriáival. Ráadásul a Facebook Hirdetéskezelő segítségével nem minden országra készíthetők becslések.

2.3. Létszámbecslési módszerek a Facebookon elérhető adatok segítségével a migrációs kutatásokban

A Facebook Hirdetéskezelője valós időben nagy, nem reprezentatív mintákból készíti el a népességbecsléseit, azaz ezek a mérőszámok becstelt és mintavételezett értékeknek tekinthetők (Zagheni et al., 2017). A profiladatok frissítései a becslésekben is megjelennek, tehát úgy tekinthetünk az oldalon elérhető adatkészletre, mint egy folyamatosan frissített népszámlálási adatra.

A Facebook Hirdetéskezelőnek ezeket a tulajdonságait először Zagheni és munkatársai (2017) használták fel az Egyesült Államokban élő migránsállomány becslésére. Megközelítésük abban volt úttörő, hogy a Facebook Hirdetéskezelő adatainak és a hagyományos adatforrásokból származó hivatalos adatoknak a kombinációjára támaszkodtak, hogy időszerű és földrajzilag részletezett becsléseket hozzanak létre a fejlett országokban. A kutatók az életkor, a nem, a származási ország és az Egyesült Államok célállamának minden egyes kombinációja esetében megvizsgálták a külföldön született egyének *American Community Survey (ACS)* által becstelt hányadának és a Facebook-felhasználókra vonatkozó megfelelő mennyiségnek a különbségét. A két becslés közötti eltérést lineáris regressziós keretrendszer segítségével modellezték, a különböző életkorcsoportokra és származási országokra vonatkozó indikátorváltozók segítségével, annak érdekében, hogy értékeljék, milyen mértékben alakultak ki mintázatok. A szerzők azt javasolták, hogy az általuk megfigyelt szabályszerűségeket a migrációs előrejelzések javítására használják fel.

A témában fontos kutatást végeztek Rampazzo és munkatársai (2021) is, akik az Egyesült Királyság területén élő európai migráns populáció méretére és demográfiai összetételére vonatkozó becslést készítettek, bayesi módszerekkel. Munkájukban a Facebook Hirdetéskezelőből kinyert adatokat és a *Labor Force Survey (LFS)* migrációs adatait kombinálták. A két adatforrás hibákkal terhelt, a modellben torzítási, fedési és definíciós hibákat is figyelembe vettek. Kutatásuk jelentősége abban áll, hogy a modelljükbe beilleszthetők új információk is, beleértve a további digitális forrásból származó (például Instagram- vagy LinkedIn-) adatokat is.

Spyratos és munkatársai (2018) is egy olyan módszert mutattak be, amellyel a Facebook Hirdetéskezelő nem reprezentatív „expat statisztikái” alapján becslhető meg a migránsállomány nagysága. Módszerük azonban egyedülálló abban,

hogyan célja nem a migrációs statisztikák reprodukálása volt, hanem az, hogy külön becsléseket állítsanak elő a kivándorlókról – éppen ezért regressziós módszert nem alkalmaztak. Kutatásukban 17 európai célországot és 82 korábbi lakóhely szerinti országot vettek figyelembe. A Facebook-felhasználók valós populációhoz viszonyított túl- vagy alulreprezentáltságának korrigálására egy együtthetót vezettek be, amely nemcsak a korábbi lakóhely szerinti országra, célországra, korcsoportra és nemre vonatkozóan egyedi, hanem figyelembe veszi a hivatalos (Eurostat, UNDESA) statisztikákat felhasználva azt is, hogy a különböző korú és nemű migránsok mennyire asszimilálódnak a célország Facebook-hálózat-használati szintjéhez (ehhez az egyes migráncsoportok Facebook-penetrációját a jelenlegi és a korábbi lakóhely szerinti országok Facebook-penetrációjának súlyozott átlagaként határozták meg).

A külföldön élő populáció becslésének ismertetett módszerei nem amiatt lényeges módszertani újítások, mert egy pillanatképet adnak a populáció méretéről és struktúrájáról. Az adatfelvétel bármikor olcsón és gyorsan megismételhető, ezáltal tetszőleges időszakonként – akár 3–6 havonta – valós idejű adatokhoz juthatunk. A longitudinális adatok érthetőbbé teszik a migrációs folyamatokat, azok dinamikáját, a migrációra ható politikai, gazdasági, társadalompolitikai tényezőket. Ezen adottságok miatt az Európa Tanács (*Gendronneau et al., 2019*) is ajánlást fogalmazott meg az adatfelvételi módszer alkalmazására vonatkozóan.

3. Az Ausztriában élő magyar nők számának becslése

A Facebook Hirdetéskezelője (<https://business.facebook.com/adsmanager/>) segítségével adatokat gyűjtöttünk a Facebook-hálózat havi aktív felhasználóinak számáról aktuális tartózkodási helyük országa, életkoruk, nemük és korábbi lakóhelyük országa alapján.

A Facebook-adatoknak a referenciaadatokkal történő összevetéséhez 2022. január 1-jén lekérdeztük a Facebook-hálózat azon női felhasználóinak számát életkor-kategóriánként (15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64 éves, 65 éves és annál idősebb), akik jelenleg Magyarországon élnek, valamint azon felhasználók számát, akik most Ausztriában élnek, és azon felhasználók számát is, akik jelenleg Ausztriában élnek, nők, és korábbi lakóhelyük Magyarország volt. Őket Ausztriában élő, magyarországi migráns nőknek tekintjük.

Referenciaként az Eurostat statisztikáit használtuk. A „DEMO_PJANGROUP – Population on 1 January by age group and sex” című tábla alapján meghatároz-

tuk a Magyarországon élő nők számát és az Ausztriában élő nők számát, a „MIGR_POP3CTB – Population on 1 January by age group, sex and country of birth” című tábla alapján pedig az Ausztriában élő, Magyarországon született nők számát 2022. január 1-jén, életkor szerinti bontásban.

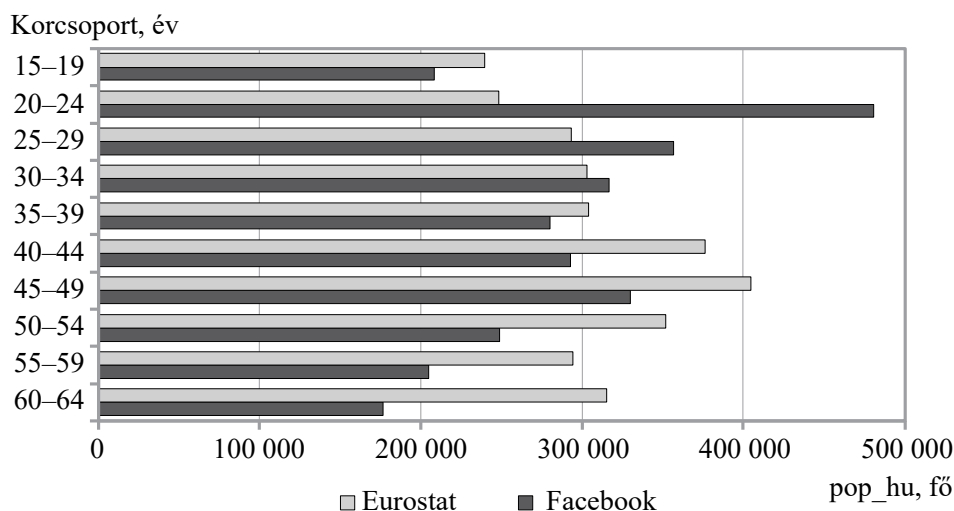
A Facebook és a referenciaadatok korcsoportonkénti eltérését Magyarország női populációjára vonatkozóan (pop_hu jelöléssel) az 1. ábrán, Ausztria női populációjára vonatkozóan (pop_au jelöléssel) a 2. ábrán, illetve az Ausztriában élő, magyar migráns nőkre vonatkozóan (migr jelöléssel) a 3. ábrán szemléltettük.

Az 1–3. ábrákon megfigyelhető, hogy a hivatalosnak tekintett Eurostat-adatok eltérése a Facebook-adatoktól kortól függően változik. E „Facebook-penetrációs ráták” hasonlósága jellemző a Magyarországon és az Ausztriában élő nők populációja esetén, azonban eltérő mintázatot mutat az Ausztriában élő, magyar migráns nők esetén.

1. ábra

Magyarország 15–64 év közötti női populációja (Eurostat) és a Facebook 15–64 év közötti Magyarországon élő női populációja korcsoportonként, 2022. január 1.

Hungary's female population aged 15–64 (Eurostat) and Facebook's female population aged 15–64 living in Hungary by age group, January 1, 2022



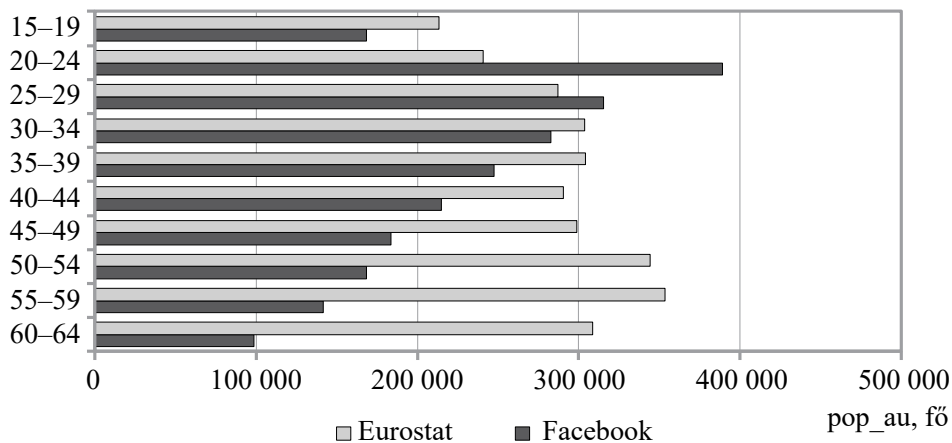
Megjegyzés: az ábrán a 65 éves és annál idősebb korosztályt nem tüntettük fel.

2. ábra

Ausztria 15–64 év közötti női populációja (Eurostat) és a Facebook 15–64 év közötti Ausztriában élő női populációja korcsoportonként, 2022. január 1.

Austria's female population aged 15–64 (Eurostat) and Facebook's female population aged 15–64 living in Austria by age group, January 1, 2022

Korcsoport, év



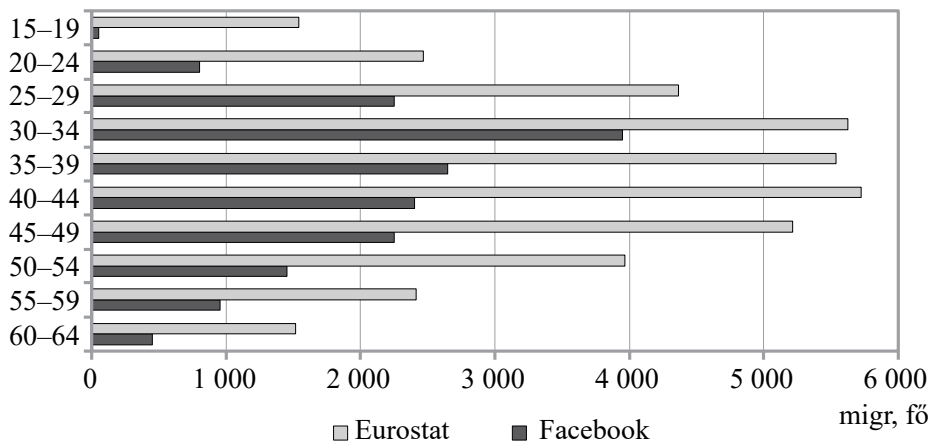
Megjegyzés: az ábrán a 65 éves és annál idősebb korosztályt nem tüntettük fel.

3. ábra

Az Ausztriában élő magyarországi migráns nők száma a Facebook és az Eurostat adatai alapján, korcsoportonként, 2022. január 1.

Number of Hungarian migrant women living in Austria based on Facebook and Eurostat data, by age group, January 1, 2022

Korcsoport, év



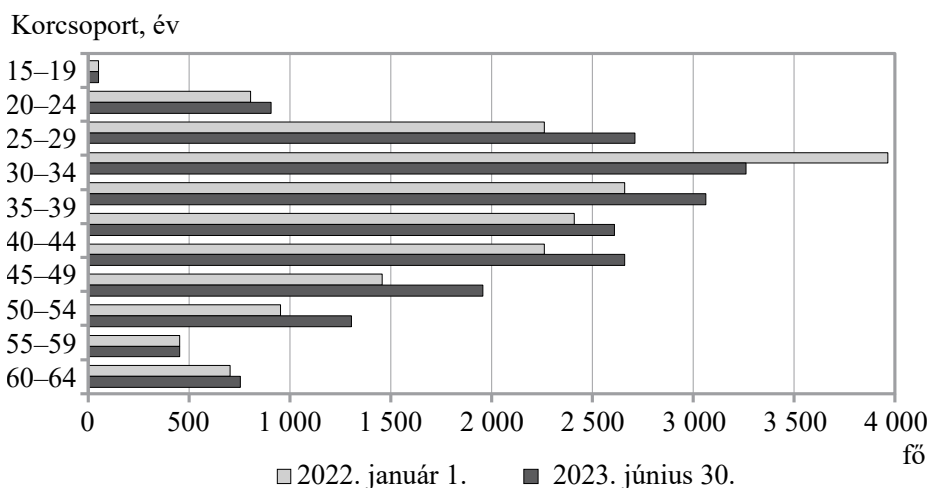
Megjegyzés: az ábrán a 65 éves és annál idősebb korosztályt nem tüntettük fel.

Másfél év elteltével, azaz 2023. június 30-án még nem állnak rendelkezésre a következő évi (2023. január 1-jei) hivatalos Eurostat-adatok, azonban a valós idejű Facebook-becslések bármikor lekérdezhetőek. A 2023. június 30-án készített Facebook-becslés eredményeit az Ausztriában élő magyar migráns nők számát mutató 4. ábrán szemléltettük. A 4. ábrán feltüntettük az adott populációra vonatkozó korábbi (2022. január 1-jei) becslésünk eredményeit is, így látathatóvá váltak a korcsoportonkénti létszámváltozások is a vizsgált időszakban.

4. ábra

**Az Ausztriában élő magyarországi migráns nők száma a Facebook adatai alapján
2022. január 1-jén és 2023. június 30-án**

*Number of Hungarian migrant women living in Austria based on Facebook data
on January 1, 2022 and June 30, 2023*



Megjegyzés: az ábrán a 65 éves és annál idősebb korosztályt nem tüntettük fel.

Ily módon két tetszőleges időpont között megbecsülhető az Ausztriában élő, magyar migráns nők számában bekövetkezett változás. Amennyiben rendelkezésre áll múltbeli becslés is – olyan időpontra vonatkozóan, amelyre elérhetőek hivatalos adatok –, úgy a Facebook-becslések kalibrálhatók a múltbeli Facebook-adatok és a múltbeli hivatalos adatok összevetésével. A kalibrációnak több módszere is ismert (*Zagheni–Weber, 2015; Zagheni et al., 2014; 2017; Spyratos et al., 2018; Rampazzo et al., 2021*), azonban e módszerek bemutatása nem célja e tanulmánynak.

Természetesen a becslés során felmerülnek különböző korlátok, amelyeket a következőkben ismertettünk.

Az adatgyűjtés során korlátot jelentett, hogy az aggregált hivatalos (Eurostat-) adatok csak előre rögzített korkategóriánként állnak rendelkezésre, valamint, hogy a tárgyévet megelőző év januárjára érhetők csak el.

A Facebook Hirdetéskezelő korlátjának tekinthető, hogy (a felhasználók személyiségi jogainak védelme érdekében, hogy megakadályozza az egyének kis csoportjának megcélzását) csak 1000 főnél nagyobb közönségre készíti el a becslést. Ennél kisebb célközönség esetén az „1000 fő alatt” eredményt kapjuk, ami megnehezíti az adatgyűjtést, ugyanis ez esetben csak kategória-összevonásokat alkalmazva, utólagos különbségképzéssel tudunk dolgozni.

Bár a Facebook által használt korkategóriák tetszőlegesen megadhatók (egész életkorokat használva), de 18 év alatti személyek csak korlátozottan, 13 év alatti személyek pedig egyáltalán nem szerepeltethetők az aggregátumokban (hiszen a korcsoportok tagjai korlátozottan, illetve nem célozhatók meg hirdetésekkel), ezenkívül a 64 év feletti korcsoport esetén további bontásra nincs lehetőség.

További mérési hibát eredményezhet, hogy egy Facebook-fiók nem feltétlenül felel meg egy személynek. Ezenkívül a földrajzi, IP-alapú helymeghatározás is hordozhat magában hibalehetőséget, ha például a felhasználó egy másik helyről proxyszervert használ a Facebook-hozzáféréshez.

Végül korlátot jelenthet az is, hogy a két adatforrásnál eltér a definíció, nevezetesen, hogy kit tekintenek migránsnak (külföldön élőknek).

4. Konklúzió

Magyarország csatlakozása az Európai Unióhoz a magyar emberek számára korábban nem látott lehetőségeket nyitott meg. A szabad mozgás és a társadalombiztosítási jogok hordozhatósága lehetővé tette, hogy az emberek az Európai Unió tagországai között szabadon mozogjanak, bárhol munkát vállaljanak. Ami az egyén szintjén lehetőség, az a nemzetállamok számára új kihívásokat jelent a gazdasági, demográfiai és szociálpolitikai tervezés területén. Az elvándorlás ellensúlyozására, negatív hatásainak a mérséklésére ezért új emberierőforrás-stratégiák kidolgozására van szükség. A tervezéshez egyrészt pontos statisztikai adatokra, a migrációs trendek mélyebb ismeretére és idősoros elemzésekre, másrészt a kibocsátó és a fogadó országok munkaerőpiaci folyamatainak folyamatos monitorozására lenne szükség. Mindezek valós idejű, pontos, megbízható és érvényes adatokat igényelnek, hogy a politika meg tudja határozni a beavatkozási pontok pontos helyét és a beavatkozás módját.

A stratégiai tervezéshez ezért a digitális demográfia új eszközeit is érdemes alkalmazni. A magyarok vándorlásáról olyan olcsó, gyors és valós idejű longitudinális adatokkal szolgálhat a Facebook, amelyek könnyen hozzáférhetők, és a leírtak alapján jól kombinálhatók az offline adatforrásokkal. A döntéshozók tehát tetszőleges időpontban, bármikor időszerű adatokhoz juthatnak. A Facebookról a létszámadatokon túl a külföldön élő magyarokról egy sor egyéb olyan, releváns adat is hozzáférhető, amelyeket a tervezéspolitikai jól használhat. Az olcsó adatokon túl természetesen léteznek más digitális adatforrások is. A Google Trends- és a mobiladatok például szintén megfelelő és pontos adatforrások, ám a Facebookhoz képest jóval költségesebbek.

Irodalom

- Abel, G. J. – Sander, N. (2014): Quantifying global international migration flows. *Science*. Vol. 343. No. 6178. pp. 1520–1522. <https://doi.org/10.1126/science.1248676>
- Barberá, P. (2016): Less is more? How demographic sample weights can improve public opinion estimates based on Twitter data. *Work Pap NYU*.
- Bijak, J. – Czaika, M. (2020): *Assessing uncertain migration futures: A typology of the unknown*. QuantMig Project Deliverable D1.1. Southampton/Krems: University of Southampton and Danube University Krems.
- Bohon, S. A. (2018): Demography in the big data revolution: Changing the culture to forge new frontiers. *Population Research and Policy Review*. Vol. 37. pp. 323–341. <https://doi.org/10.1007/s11113-018-9464-6>, <https://www.jstor.org/stable/45179508>
- Cesare, N. – Spiro, E. – Lee, H. (2015): Self-presentation and information disclosure on Twitter: Understanding patterns and mechanisms along demographic lines. *Paper presented at the annual meeting of the Population Association of America*, San Diego, CA.
- Cesare, N. – Lee, H. – McCormick, T. – Spiro, E. – Zagheni, E. (2018): Promises and pitfalls of using digital traces for demographic research. *Demography*. Vol. 55. No. 5. pp. 1979–1999. <https://doi.org/10.1007/s13524-018-0715-2>
- Engbersen, G. (2018): Liquid migration and its consequences for local integration policies. In: Scholten, P. – Van Ostaijen, M. (eds.): *Between mobility and migration: The multi-level governance of intra-European movement*. Cham. Springer International Publishing AG. pp. 63–76. https://doi.org/10.1007/978-3-319-77991-1_4
- Engbersen, G. – Snel, E. (2013): Liquid migration: Dynamic and fluid patterns of post-accession migration flows. In: Glorius, B. – Grabowska, I. – Kuvik, A. (eds.): *Mobility in transition: Migration patterns after EU enlargement*. Amsterdam. University Press. pp. 21–40. <https://doi.org/10.1017/9789048515493.002>, <https://www.jstor.org/stable/j.ctt46mwhx.4>
- Facebook (2022): *Advanced Targeting*. <https://developers.facebook.com/docs/marketing-api/audiences/reference/advanced-targeting>. (letöltés dátuma: 2022. augusztus)
- Gendronneau, C. – Wiśniowski, A. – Yildiz, D. – Zagheni, E. – Fiorio, L. – Hsiao, Y. – Hoorens, S. (2019): *Measuring labour mobility and migration using big data: Exploring the potential of social-media data for measuring EU mobility flows and stocks of EU movers*. Publications Office of the European Union.

- Golovics J. (2019): Migrációs statisztikák: kihívások és dilemmák. *Statisztikai Szemle*. 97. évf. 10. sz. 933–948. o. <https://doi.org/10.20311/stat2019.10.hu0933>
- Hárs Á. – Sik E. (2009): Hungary. In: H. Fassmann – U. Reeger – W. Sievers (eds.): *Statistics and Reality: Concepts and Measurements of Migration in Europe*. Amsterdam University Press. pp. 169–194. <http://www.jstor.org/stable/j.ctt46n2qg.13>
- Herdağdelen, A. – Marelli, M. (2017): Social media and language processing: How Facebook and Twitter provide the best frequency estimates for studying word recognition. *Cognitive science*. Vol. 41. No. 4. pp. 976–995. <https://doi.org/10.1111/cogs.12392>
- Herdağdelen, A. – State, B. – Adamic, L. – Mason, W. (2016): The social ties of immigrant communities in the United States. *Proceedings of the 8th ACM Conference on Web Science*. pp. 78–84. <https://doi.org/10.1145/2908131.2908163>
- Kádár, B. – Gede, M. (2021): Tourism flows in large-scale destination systems. *Annals of Tourism Research*. Vol. 87. 103113. <https://doi.org/10.1016/j.annals.2020.103113>
- Kmetty, Z. – Bozsonyi, K. (2022): Identifying depression-related behavior on Facebook — an experimental study. *Social Sciences*. Vol. 11. No. 3. p. 135. <https://doi.org/10.3390/socsci11030135>
- Kmetty, Z. – Németh, R. (2022): Which is your favorite music genre? A validity comparison of Facebook data and survey data. *Bulletin of Sociological Methodology*. Vol. 154. No. 1. pp. 82–104. <https://doi.org/10.1177/07591063211061754>
- Latour, B. (2007): Beware, your imagination leaves digital traces. *Times Higher Literary Supplement*. Vol. 6. No. 4. pp. 129–131.
- Lewis, S. C. – Zamith, R. – Hermida, A. (2013): Content analysis in an era of big data: A hybrid approach to computational and manual methods. *Journal of broadcasting & electronic media*. Vol. 57. No. 1. pp. 34–52. <https://doi.org/10.1080/08838151.2012.761702>
- Ligeti A. S. (2019): Cirkuláris migráció Magyarországon. *Statisztikai Szemle*. 97. évf. 4. sz. 327–346. o. <https://doi.org/10.20311/stat2019.4.hu0327>
- Massicotte, P. – Eddelbuettel, D. (2020): gtrendsR: *Perform and display Google Trends Queries*. R package version.
- Mohebbi, M. – Vanderkam, D. – Kodysh, J. – Schonberger, R. – Choi, H. – Kumar, S. (2011): *Google correlate whitepaper*.
- Pöttschke, S. – Braun, M. (2017): Migrant sampling using Facebook advertisements: A case study of polish migrants in four European countries. *Social Science Computer Review*. Vol. 35. No. 5. pp. 633–653. <https://doi.org/10.1177/0894439316666262>
- Rampazzo, F. – Bijak, J. – Vitali, A. – Weber, I. – Zagheni, E. (2021): A framework for estimating migrant stocks using digital traces and survey data: An application in the United Kingdom. *Demography*. Vol. 58. No. 6. pp. 2193–2218. <https://doi.org/10.1215/00703370-9578562>
- Sik E. (2003): A migrációs potenciál kutatásának alapfogalmai. In: Örkény A. (szerk.): *Menni vagy maradni? Kedvezménytörvény és migrációs várakozások*. MTA Kisebbségkutató Intézet, Budapest. 15–18. o.
- Sik E. (2013): Migráció a mai Magyarországon. Bevezetés. *Magyar Tudomány*. 174. évf. 3. sz. 242–243. o.
- Sik E. (2018). Migrációs burok és migrációs potenciál. In: Siskáné Szilasi B. – Halász L. (szerk.): *Boldogulni itthon vagy külföldön? ME Földrajz-Geoinformatika Intézet, Budapest*. 105–126. o.
- Spyratos, S. – Vespe, M. – Natale, F. – Weber, I. – Zagheni, E. – Rango, M. (2018): *Migration data using social media: A European perspective*.

- Spyratos, S. – Vespe, M. – Natale, F. – Weber, I. – Zagheni, E. – Rango, M. (2019): Quantifying international human mobility patterns using Facebook Network data. *PloS one*. Vol. 14. No. 10. e0224134. <https://doi.org/10.1371/journal.pone.0224134>
- U.S. SEC. (2020): *Facebook, Inc. 2019 annual report, form 10-K*.
<https://sec.report/Document/0001326801-20-000013/fb-12312019x10k.htm>
- Willekens, F. – Massey, D. – Raymer, J. – Beauchemin, C. (2016): International migration under the microscope. *Science*. Vol. 352. No. 6288. pp. 897–899.
<https://doi.org/10.1126/science.aaf6545> <https://www.jstor.org/stable/24744591>
- Zagheni, E. – Garimella, V. R. K. – Weber, I. – State, B. (2014): Inferring international and internal migration patterns from twitter data. *Proceedings of the 23rd international conference on world wide web*. pp. 439–444. <https://doi.org/10.1145/2567948.2576930>
- Zagheni, E. – Weber, I. (2012): You are where you e-mail: using e-mail data to estimate international migration rates. *Proceedings of the 4th annual ACM web science conference*. pp. 348–351.
<https://doi.org/10.1145/2380718.2380764>
- Zagheni, E. – Weber, I. (2015): Demographic research with non-representative internet data. *International Journal of Manpower*. Vol. 36. No. 1. pp. 13–25.
<https://doi.org/10.1108/IJM-12-2014-0261>
- Zagheni, E. – Weber, I. – Gummadi, K. (2017): Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*. pp. 721–734.
<https://doi.org/10.1111/padr.12102>